

Salamanova, Evdokiya (2012) *Computational studies of mechanisms, spectra and conformations in biological systems.*

PhD thesis

<http://theses.gla.ac.uk/3264/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

**COMPUTATIONAL STUDIES OF
MECHANISMS, SPECTRA AND
CONFORMATIONS IN BIOLOGICAL
SYSTEMS**

By

Evdokiya K. Salamanova

Submitted in fulfilment of the requirements for the Degree of
Doctor of Philosophy

School of Chemistry
University of Glasgow
April 2011

Abstract

This thesis represents computational studies of the mechanisms of chlorination in the FAD-dependent enzyme PrnA using QM, MD and QM/MM methods. It also contains Time-Dependent Density Functional Theory (TDDFT) spectra and conformational analysis of the fluoro-imine derivatives of pyridoxal 5'-phosphate (PLP).

Chapter 1 gives a brief introduction of the theoretical methods used in the thesis and an explanation of their implementation in Computational Chemistry. Chapter 2 presents an overview of the variety of halogenated natural products and halogenating enzymes. This chapter takes a closer look at the group of the FAD-dependent enzymes and in particular at PrnA. All the feasible mechanisms of regioselective chlorination of tryptophan in PrnA are thoroughly discussed. In Chapter 3 model reactions for the chlorination of the substrate tryptophan at the DFT level are investigated. Two different chlorinating agents are proposed and the reaction energies and barriers for the halogenating reactions are calculated. A simple comparison of the chlorinating agents at the QM level without the enzyme is given. Chapters 4 and 5 are devoted to MD simulations on PrnA for equilibration of the solvated protein and QM/MM modelling with the AM1/AMBER03 and the DFT/AMBER03 methods. The QM/MM methods gave the opportunity of studying the reaction mechanisms with the account of the environment of the enzyme. All the steps along the mechanisms of chlorination with the possible chlorinating agents were investigated.

In chapter 6 TDDFT has been used to calculate the electronic spectra of PLP and its derivatives. PLP is co-factor in a diversity of enzymes with different biological function. In the alanine racemase it takes part in the racemisation of L-alanine to D-alanine. The latter is a building block of the bacterial cell wall. A novel method of blocking the function of the enzyme is the use of the β,β,β -trifluoroalanine inhibitor which forms an external aldimine with PLP. A mechanism of the β,β,β -trifluoroalanine inhibitor activity was proposed by Faraci and Walsh.¹ The spectra at all the steps along the proposed mechanism with intermediates in different protonation states were calculated.

Chapter 7 focusses on the examination of the conformations of the PLP derivatives and the energetics of these compounds. The preference of *gauche* conformation in β -fluoroimines was studied. In this chapter are presented empirical (X-ray) and theoretical (DFT) evidence for the *gauche* effect in β -fluoroimines. NCCF bond rotational profiles of acetaldehyde-derived imines and β -fluoropyridoximine at the DFT level were analysed. NBO analysis of the investigated compounds proved the Dunathan's stereoelectronic hypothesis.

Content:

Abstract.....	ii
List of figures.....	vii
List of tables.....	xiv
Acknowledgments.....	xvi
Author's declaration.....	xvii
Chapter 1 Methods.....	1
1.1 Quantum Chemistry Theory	1
1.1.1 Electronic structure and wavefunction.....	1
1.1.2 Born-Oppenheimer approximation	2
1.1.3 The Slater determinant.....	4
1.1.4 Density Functional Theory (DFT)	5
1.1.4.1 Hohenberg-Kohn theorems.....	5
1.1.4.2 Kohn-Sham approach.....	8
1.1.4.3 Exchange-correlation functionals	9
1.1.4.3.1 The Local Density Approximation (LDA)	9
1.1.4.3.2 The Generalised Gradient Approximation (GGA)	10
1.1.4.3.3 Hybrid functionals	11
1.1.5 Time-Dependent Density Functional Theory (TDDFT).....	12
1.1.6 Semi-empirical Methods.....	16
1.1.7 NDDO formalism.....	17
1.1.7.1.1 AM1	20
1.2 Molecular Mechanics (Empirical force fields)	21
1.3 QM/MM methods	24
1.4 Molecular dynamics.....	27
1.4.1 Simulation protocol.....	30
1.4.2 Statistical mechanics.....	31
1.4.2.1 Microcanonical (NVE) ensemble	31
1.4.2.2 Canonical (NVT) ensemble and Boltzmann factor.....	32
1.4.2.3 Isothermal-isobatic (NPT) ensemble	33
1.4.2.4 Temperature coupling	33
1.4.2.4.1 Berendsen temperature coupling.....	33
1.4.2.4.2 Nosé-Hoover thermostat	34
1.4.2.4.3 Pressure coupling	34
1.4.2.4.4 Berendsen barostat.....	35
1.4.2.4.5 Parrinello-Rahman thermostat	35
1.4.3 Applications of Molecular Dynamics	35
Chapter 2 Enzymatic halogenation	37
2.1 Halogenation in nature and halogenated natural products.....	37
2.2 Halogenation strategies.....	41

2.2.1 Halogenation by Hypohalite, $\text{XO}^- (\text{X}^+)$	41
2.2.1.1 Haem-Dependent haloperoxidases.....	41
2.2.1.2 Vanadium-dependent haloperoxidases	43
2.2.1.3 Flavin-dependent halogenases	44
2.2.1.3.1 PyrH.....	46
2.2.1.3.2 RebH.....	47
2.2.1.3.3 PrnA.....	49
2.2.2 Halogenation via Halogen, X^\cdot	54
2.2.3 Halogenation via Halide, X^-	55
2.3 Electrophilic aromatic substitution	57
2.4 Possible electrophiles involved in the tryptophan chlorination at position C7..	58
Chapter 3 QM calculations on PrnA.....	60
3.1 Introduction.....	60
3.2 Computational methodology.....	62
3.2.1 Computational models and augmented basis set	62
3.2.2 COSMO (Conductor-like Screening Model) calculations.....	62
3.2.3 Transition state search.....	65
3.3 Results and Discussion	66
3.3.1 Stabilities of chlorinated indoles.....	66
3.3.2 Possible electrophiles involved in the indole chlorination	67
3.3.3 Energy profiles for electrophilic chlorination of indole	69
3.3.4 Transition state search with <i>ChemShell</i>	73
3.3.5 Pre-Reaction Complexes.....	75
3.4 Conclusions.....	76
Chapter 4 Classical Molecular Dynamics simulations on PrnA monomer.....	78
4.1 Introduction.....	78
4.2 Computational methodology.....	79
4.2.1 Residues building.....	80
4.3 Results and Discussion	83
4.3.1 MD simulations with FADHOOH in the PrnA monomer	83
4.3.1.1 Preparation of the initial structure for MD simulations and simulation parameters	83
4.3.1.2 Total RMSD from X-ray structures	88
4.3.1.3 RMSD values of the protein residues	90
4.3.1.4 RMSD values of the ligands	92
4.3.1.5 Deviation in the Lys78 residue	93
4.3.1.6 The chloride binding site	93
4.3.1.6.1 LYN	93
4.3.1.6.2 LYP.....	95
4.3.2 MD simulations with FAD in the PrnA monomer and dimer.....	98
4.4 Conclusions.....	100
Chapter 5 QM/MM modelling on PrnA dimer	102
5.1 Introduction.....	102
5.2 Computational methodology.....	104
5.2.1 Validation of O-O torsional parameters.....	105
5.2.2 Preparation of the system for the QM/MM simulations	107
5.3 Results and Discussion	109
5.3.1 Nucleophilic attack of the chloride on FADHOOH	109

5.3.2 Water networks and protonation state of Lys78	112
5.3.3 Attack of the hypochlorous acid on the Lyn78 residue	114
5.3.4 Formation of the Lyn78-chloramine intermediate	116
5.3.5 Formation of the Tryptophan σ - complex	119
5.3.5.1 (Chloroammonio)lysine78 as chlorinating agent	119
5.3.5.2 Hydrogen-bonded hypochlorous acid as chlorinating agent	121
5.4 Conclusions	123
Chapter 6 Pyridoxal derivatives: Structural and spectroscopic investigation.....	128
6.1 Introduction	128
6.2 Computational methodology	133
6.3 Results and Discussion	133
6.3.1 Validation spectra	134
6.3.2 External aldimines' structures and spectra	141
6.3.3 Structures and spectra of quinonoids (1a)	144
6.3.4 Structures and spectra of β -difluoro- α,β -unsaturated imine	147
6.3.5 Structures and spectra of the enzyme-pyridoxal aldimine complexes	148
6.3.6 Structures and spectra of (3a)	150
6.3.7 Structures and spectra of the inactive complex (4)	152
6.4 Conclusions	153
Chapter 7 Theoretical and X-ray crys-tallographic evidence of a fluorine-imine gauche effect: an addendum to Dunathan's hypothesis	156
7.1 Introduction	156
7.1.1 Experimental structures in the solid state	157
7.2 Computational methodology	159
7.3 Results and Discussion	159
7.3.1 Computational conformational analysis	159
7.3.1.1 Conformational preferences of β -fluoroimines	159
7.3.1.2 Natural Bond Orbital analyses	162
7.3.1.3 Conformational preferences of β -fluoropyrido- ximine	163
7.3.1.4 C-F activation in β -fluoropyridoximine	168
7.4 Conclusions	170
 References	 171
Appendices	177

List of Figures

Figure 2-1. Chlorinated dioxins and benzofurans.....	37
Figure 2-2. Chlorinated compounds isolated from lichen (1) and fungi.....	38
Figure 2-3. Chlorinated antitumor agent Cryptophycin-1 discovered in the alga	38
Figure 2-4. Organohalogen, produced by bacteria with HIV-inhibitor (4) and antibiotic (5) activities ³⁹	39
Figure 2-5. The organohalogen telfairine (6) and a monoterpene (7) isolated from inmarine organisms ³⁹	39
Figure 2-6. Organohalogen in human: the thyroid hormone thyroxine (8) and bromoester found in the cerebrospinal fluid and responsible for the REM sleep (9) ³⁹	40
Figure 2-7. a) caldariomycin; b) monochlorodimedone; c) 2-chloro-1,3-cyclopentanedione ⁴²	42
Figure 2-8. Formation hypochlorous acid by haem-dependent chloroperoxidases ^{45,52}	43
Figure 2-9. Formation of vanadium bound hypobromite by vanadium-dependent haloperoxidase ^{45,52}	43
Figure 2-10. Enzymes from the flavin-dependent halogenase family, catalysing regioselective chlorinations ^{49,63}	46
Figure 2-11. PyrH dimer structure downloaded from PDB (2WET.pdb) ⁶⁶	47
Figure 2-12. Crystal structure of the active centre of PyrH with Cl ⁻ , FAD and tryptophan bound, downloaded from PDB (2WET.pdb) ⁶⁶	47
Figure 2-13. RebH structure with bound FAD and Cl ⁻ , downloaded from PDB (2OAL.pdb) ⁶⁶	48
Figure 2-14. Formation of 7-chlorotryptophan as a first step of the biosynthesis of rebeccamycin by RebH ⁶⁸	49
Figure 2-15. Stereo pictures of the PrnA monomer. A) The N-terminus of the enzyme is coloured in blue and the C-terminus in red, the chloride is a pink sphere, FAD and 7-chlorotryptophan are represented in sticks. B) The FAD and chloride (B1) and the substrate-binding (B2) sites of the enzyme are shown in sticks. The hydrogen bonds are shown as dotted lines ⁶⁴	50
Figure 2-16. Biosynthesis of the natural antibiotic pyrrolnitrin in <i>P. fluorescens</i> ⁴²	51

Figure 2-17. Formation of 7-chlorotryptophan: A) Chlorination by the free HOCl; B) Walsh and co-workers mechanism of chlorination through a formation of chloramine with the participation of Lysine78; ⁶⁸ C) Naismith et <i>al.</i> mechanism of chlorination with hydrogen bound HOCl to the Lysine78 residue ⁶⁴	53
Figure 2-18. Non-haem dependent halogenase in which previous to the halogenation the substrate has to be bound to the thiolation domain of the SyrB1 carrier ⁵⁰	54
Figure 2-19. Mechanism of halogenation by X ⁻ of non-haeme iron O ₂ ⁻ and α-ketoglutarate-dependent halogenases ⁵²	55
Figure 2-20. Fluorinase mechanism of halogenation involving transformation of F ⁻ and SAM to 5'-FDA and L-methionine and further metabolism of 5'-FDA to fluoroacetate and 4-fluorothreonine ⁵⁰	56
Figure 2-21. Mechanism of electrophilic aromatic substitution	57
Figure 2-22. Formation of hypochlorous acid	58
Figure 2-23. Hydrogen bond between the hypochlorous acid and the Lys78-residue.	58
Figure 2-24. Reaction between protonated Lys78-residue and the hypochlorous acid	58
Figure 2-25. Reaction between the hypochlorous acid non-protonated Lys78	59
Figure 3-1. Active centre in the crystal structure of PrnA	60
Figure 3-2. Formation of 7-Chlorotryptophan from the attack of different halogenating agents on tryptophan	61
Figure 3-3. The model compounds used for the reaction of chlorination of tryptophan to 7-chlorotryptophan occurring in the active centre of the PrnA	61
Figure 3-4 The investigated positions for the indole chlorination.....	66
Figure 3-5. Formation of hypochlorous acid	68
Figure 3-6. Reaction between hypochlorous acid and non-protonated Lys78.....	68
Figure 3-7. Reaction scheme for the formation of protonated hypochlorous acid	68
Figure 3-8. Formation of hypochlorous acid	68
Figure 3-9. Formation of a chloramine	68
Figure 3-10. Formation of (chloroamino)methane	68
Figure 3-11. Formation of protonated hypochlorous acid	69
Figure 3-12. Potential energies of the σ-complexes of 7-chloroindole and products of hypochlorous acid + indole and (chloroammonio)methane + indole reactions, relative to the reactants at the TPSS/def2-TZVP+//IEFPCM(ε _r = 78.4) level of QM theory	70

Figure 3-13. Energy profiles with optimised geometries at each point at the TPSS/def2-TZVP+ level with constraints of: A)(chloroammonio)methane + indole and B) hypochlorous acid + indole reactions.....	71
Figure 3-14. Energy profiles with optimised geometries at each point at the TPSS/def2-TZVP+ level with the restraints of: A) (chloroammonio)methane + indole and B) hypochlorous acid + indole reactions. The profiles start on the right side with the reactants and move along the curve in the direction of the products on the left side	72
Figure 3-15. Starting structures and numbering of the atoms in the reactants: A) hypochlorous acid + indole; B) (chloroammonio)methane + indole	73
Figure 3-16. Transition state of hypochlorous acid + indole reaction at the TPSS/def2-TZVP+ level in IEFPCM continuum water	74
Figure 3-17. Transition state of (chloroammonio)methane + indole reaction at the TPSS/def2-TZVP+ level in IEFPCM continuum water	74
Figure 3-18. Pre-reaction complex of (chloroammonio)methane and indole at the TPSS/def2-TZVP+ level in IEFPCM continuum water	76
Figure 3-19. Pre-reaction complex of hypochlorous acid and indole at the TPSS/def2-TZVP+ level in IEFPCM continuum water	76
Figure 4-1. Optimised FAD structure at the HF/6-31G** level	81
Figure 4-2. Optimised ZTRP structure at the HF/6-31G**//IEFPCM($\epsilon_r = 4.24$) level	81
Figure 4-3. FADHOOH optimised structure at the HF/6-31G** level with 105.1° value of the C-O-O-H torsion	82
Figure 4-4. FADHOOH optimised structure at the HF/6-31G** level with -78.9° value of the C-O-O-H torsion	82
Figure 4-5. X-ray and initial LYP-system structure after adding hydrogens and building the peroxy group of FADHOOH	86
Figure 4-6. Minimised LYN and LYP structures with steepest descent using ffAMBER03. Hydrogen-bonding interactions of the chloride (orange).....	87
Figure 4-7. Lys78 structures: A) X-ray structure; B) and C) Initial LYN78 and LYP78 structures; D) and E) Minimised LYN78 and LYP78 structures using ffAMBER03.....	88
Figure 4-8. Total RMSD for all the protein atoms except the hydrogen from the production run.....	89
Figure 4-9. RMSD's averaged over the final 5ns of LYN and LYP protein heavy atoms from the X-ray structure per residue	90
Figure 4-10. Positions of protein residues with high RMSDs. A) Deviation of the last snapshot of the simulated PrnA monomer (cyan) from the X-ray-structure (gray). The deviated residues are shown in green for the simulated PrnA, and in magenta	

for the crystal structure; B) PrnA dimer. The deviated residues are shown in red	92
Figure 4-11. Chloride surroundings in the X-ray structure (A), after the initial minimisation(B) and after the production run of the MD simulation with LYN-last snapshot (C).....	94
Figure 4-12. Chloride-water distances in the chloride binding site of LYN78 in the last 5ns of the simulation	95
Figure 4-13. Distribution of the distances between the chloride and the oxygen atoms of water molecules 536, 563 and 700 in the LYN system	95
Figure 4-14. Chloride surroundings in the X-ray structure (A), after the initial minimisation (B) and after the production run of the MD simulation with LYP-last snapshot (C).....	96
Figure 4-15. Chloride-waters distances in the chloride binding site of LYP78, in the last 5ns of the simulation	97
Figure 4-16. Distribution of the distances between the chloride and the oxygen atoms of water molecules 536, 563 and 700 in the LYP system.....	97
Figure 4-17. Comparison between Cl-O distances in the chloride environment	98
Figure 4-18. Position of the chloride through the simulation in LYN78. The initial position is shown in green and every 100 th frame (every 100 ps) is in orange....	98
Figure 5-1. QM regions in the optimised initial structures at the AM1/AMBER03 level along the mechanism of the chlorination in PrnA.....	103
Figure 5-2. Mechanisms of the reactions, subject to QM/MM studies.....	104
Figure 5-3. Structure of 5-methyl-4 α -peroxy-10-H-isoalloxazine	105
Figure 5-4. Scans with B3LYP, AM1 and PM3	106
Figure 5-5. PrnA dimer setup system	108
Figure 5-6. Optimised QM/MM FADHOOH and chloride reactants.....	109
Figure 5-7. Optimised transition state structure for the attack of Cl519 on FADHOOH at the AM1/AMBER03 level	110
Figure 5-8. QM/MM optimised structures of the products of HOCl formation	111
Figure 5-9. Water networks connecting Lyp78 and Glu345, after an optimisation at the AM1/AMBER03 level	113
Figure 5-10. Net reaction of proton transfer between Lyp78 and Glu345.....	113
Figure 5-11. Adduct structures of the hypochlorous acid attacking Lyn78 with TPSS/def2-TZVP+/AMBER03	115
Figure 5-12. Product A of the HOCl attack on Lyn78 with TPSS/def2-TZVP+/AMBER03	116

Figure 5-13. Optimised QM/MM structure in PrnA of the (chloroamino)lysine product A with TPSS/def2-TZVP+/AMBER03	117
Figure 5-14. Optimised product B of (chloroammonio)lysine in PrnA with AM1/AMBER03	118
Figure 5-15. Optimised product B of (chloroammonio)lysine in the enzyme with TPSS/def2-TZVP+/AMBER03	118
Figure 5-16. Starting QM/MM structure of the electrophilic attack on tryptophan at the DFT(TPSS/def2-TZVP+)/AMBER03 level	119
Figure 5-17. AM1/AMBER03 optimised structures of A) the (chloroammonio)lysine-tryptophan complex; B) the transition state of the (chloroammonio)lysine + tryptophan reaction and C) the products (the σ -complex of 7-chloroindole and Lyn78).....	120
Figure 5-18. QM/MM Optimised Lyp78, HOCl and tryptophan and chloride reactants	121
Figure 5-19. Optimised transition state structure of the attack of HOCl on tryptophan at the AM1/AMBER03 level	122
Figure 5-20. AM1/AMBER03 optimised structures of the products of the attack of HOCl on tryptophan.....	123
Figure 6-1. Vitamin B ₆ (pyridoxine) and its derivatives.....	128
Figure 6-2. Reactions with PLP as a co-factor in variety of enzymes ¹²⁹	129
Figure 6-3. Mechanism of β -Replacement for tryptophan synthase ¹²⁹	130
Figure 6-4. Resonance structures of the deprotonated external aldimine	130
Figure 6-5. Spectrum of a mixture of 4.96 x 10 ⁻⁵ M PLP and 0.02 M diethyl aminomalonate at pH 6.2. The time in minutes is given by each curve ¹³⁰	131
Figure 6-6. Mechanism of inhibition of alanine racemase involving β,β -trifluoroalanine ¹	132
Figure 6-7. Calculated absorption spectrum of H ₃ P at the M06-2X/6-311+G(2df, p) level in gas phase	136
Figure 6-8. Molecular orbitals' representations of H ₃ P generated at the M05-2X/6-311+G(2df, p) level in gas phase: A) <u>Excited state 2</u> : λ = 303.74 nm, f= 0.19, E= 4.08 eV, 64 \rightarrow 65 orbitals transition, CI= 0.70; B) <u>Excited state 6</u> : λ = 199 nm, f=0.15, E= 6.23 eV, 64 \rightarrow 66 orbitals transition, CI= 0.50; C) <u>Excited state 7</u> : λ = 196 nm, f= 0.12, E= 6.34 eV, 59 \rightarrow 65 orbitals transition, CI= 0.25	137
Figure 6-9. A) Resonance structures of H ₂ P \pm ; B) A final structure after an optimisation of H ₂ P \pm at the M06-2X/6-311+G(2df, p) level in gas phase	138

Figure 6-10. Experimentally measured spectrum of the most anionic form of PLP; the major band corresponds to the aldehyde and the small bands are the covalent hydrate ¹³⁶	139
Figure 6-11. Calculated spectra of PLP or pyridoxal: A) Optimised PLP in gas phase with spectrum calculated in gas phase; B) Optimised pyridoxal in gas phase with spectrum calculated in IEFPCM water; C) Optimised pyridoxal in IEFPCM water and spectrum in IEFPCM water	139
Figure 6-12. External aldimines. Tautomeric forms of the first intermediates of the mechanism of inhibition of alanine racemase, proposed by W. S. Faraci and C. T. Walsh ¹ in their possible protonation states depending on the pH	141
Figure 6-13. Calculated spectra of the external aldimines with M06-2X/6-311+G(2df, p) in IEFPCM continuum water	142
Figure 6-14. Orbitals, generated at the M05-2X/6-311+G(2df, p) level in gas phase, corresponding to the first two peaks in the B1 (1) spectrum: A) <u>Excited State 2</u> : E= 3.3568 eV, λ = 369.35 nm, f= 0.2840, 75 -> 76 orbitals transition, CI= 0.69921; B) <u>Excited State 4</u> : E= 4.9073 eV, λ = 252.65 nm, f= 0.1760, 73 -> 76 orbitals transition, CI= 0.68291	143
Figure 6-15. Tautomerisation of B1 (1a) to quinonoid structure B2 (1a) optimised with M06-2X/6-311G(2df, p) in gas phase	144
Figure 6-16. Structures of deprotonated external aldimines (1a) intermediates (second step in the mechanism of the alanine racemase inhibition) ¹	144
Figure 6-17. Calculated spectra of the (1a) compounds at the M06-2X/6-311G(2df, p) in IEFPCM continuum water	145
Figure 6-18. Orbital at the M05-2X/6-311+G(2df, p) level in gas phase of the first excited state in the quinonoid form (1a): <u>Excited State1</u> : E= 3.0614 eV, λ = 405 nm, f= 0.8761, 75 -> 76 orbitals transition, CI= 0.70307	145
Figure 6-19. Reaction energies of the protonation/deprotonation against $\text{CH}_3\text{NH}_3^+/\text{CH}_3\text{NH}_2$ of aldimines and quinonoids of A2 and B2 at the M06-2X/6-311G(2df, p) level in vacuum	146
Figure 6-20. Structures of β -difluoro- α,β -unsaturated imine intermediates (compound 2) optimised with M06-2X/6-311G(2df, p) in vacuum (third step mechanism proposed by Wash <i>et al.</i> for the alanine racemase inhibition) ¹	147
Figure 6-21. Calculated spectra of the β -difluoro- α,β -unsaturated imines in IEFPCM continuum water using M06-2X/6-311G(2df, p)	147
Figure 6-22. Addition products from the reaction between the enzyme residue and the pyridoxal-inhibitor complex in the Walsh <i>et al.</i> mechanism ¹	148
Figure 6-23. Calculated spectra of the (3) structures in different protonation states in IEFPCM continuum water using M06-2X/6-311G(2df, p)	149

Figure 6-24. Mulliken atomic charges of the structures A2 (3) and C (3) before and after a nucleophilic attack	149
Figure 6-25. Resonance structures of B2 (3)	150
Figure 6-26. Bond distances in B2 (3) before (A) and after the optimisation (B) at the M06-2X/6-311G(2df, p) level in gas phase	150
Figure 6-27. Structures of (3a) compounds, deprotonated intermediates with B2 (3a) and C (3a) quinonoid structures	151
Figure 6-28. Calculated spectra of (3a) species with M06-2X/6-311G(2df, p) in IEFPCM continuum water	151
Figure 6-29. Final products (4) of the mechanism of inhibition of the alanine racamase by β,β,β -trifluoroalanine ¹	152
Figure 6-30. Resonance forms of B2 (4) product	152
Figure 6-31. Calculated spectra of (4) compounds at the M06-2X/6-311G(2df, p) level in IEFPCM continuum water	153
Figure 7-1. The Dunathan Hypothesis to correlate reaction specificity and conformation in PLP-dependent enzymes ¹⁴¹	157
Figure 7-2 Newman projections of β -fluoroimines 6 , 7 and 8	158
Figure 7-3. Bond rotation profiles of acetaldehyde-derived imines (F vs. H), calculated in vacuum at the M06-2X/6-311G(2df, p) level.....	160
Figure 7-4. Bond rotation profiles of β -fluoropyridoximine tautomers 8a and 8b , calculated in water at the M06-2X/6-31+G(d, p) level. Relative energies of stationary points using the larger 6-311G(2df, p) basis set differ by <1 kJ mol ⁻¹ (data not shown).....	166
Figure 7-5. Bond rotation profiles of the quinoid β -fluoropyridoximine tautomers (<i>E</i>)- and (<i>Z</i>)- 8c , calculated in water at the M06-2X/6-31+G(d, p) level. Data for both isomers are plotted relative to the lowest stationary point at this level of theory, (<i>-ac</i>)-(<i>E</i>)- 8c . Full black symbols show conformer energies obtained with the larger 6-311G(2df, p) basis set, relative to lowest stationary point at that level, which is (<i>syn</i>)-(<i>E</i>)- 8c	167
Figure 7-6. Unfavourable interactions responsible for the highest NCCF rotational barrier in (<i>E</i>)- and (<i>Z</i>)- 8c , respectively.	168
Figure 7-7. Variation of the C–F bond length with NCCF torsion in (<i>E</i>)- 8c , calculated using M06-2X/6-31+G(d, p) in IEFPCM continuum water.	169
Figure 7-8. Illustration of the stereoelectronic C–F activation in (<i>E</i>)- 8c at $\phi_{\text{NCCF}} = 90^\circ$. Shown are the occupied π_{NC} and the vacant σ^*_{CF} NBOs.	169

List of Tables

Table 2-1. FAD-dependent halogenases	45
Table 3-1. Exponents added to the diffused functions in the def2-TZVP+ basis set ..	62
Table 3-2 Calculated and experimental solvation free energies and deviations	63
Table 3-3. Calculated and experimental solvation energies with different radii	64
Table 3-4. Default and optimised solvation radii used in the calculations	65
Table 3-5. Stability of the chlorinated compounds relative to 7-chloroindole (C7) at the TPSS/def2-TZVP+ level in IEFPCM water	67
Table 3-6. Distances in reactants, transition states and products of the chlorinating process, with two different chlorinating agents	74
Table 4-1. pKa values for the important residues	79
Table 4-2. Volumes of box types for the PrnA monomer.....	83
Table 4-3. Description of LYN and LYP systems	84
Table 4-4. Simulation parameters	85
Table 4-5. Deviation of all protein atoms except hydrogen of LYN and LYP systems from the X-ray structure, during the final 5 ns of the simulations.....	90
Table 4-6. RMSD values per protein residue for the protein atoms except hydrogen averaged over the final 5 ns of the simulations.....	91
Table 4-7. RMSD values of the ligands over the final run trajectory	93
Table 4-8. RMSD values of the lysine 78 residue in LYP and LYN from the X-ray structure.....	93
Table 4-9. Average RMSD values of the residues present in the dimer interface in various molecules. (Chain A: Residues 1 to 521 Chain B: Residues 522 to 1042)	99
Table 4-10. RMSD values of ligands in monomer and dimer system	100
Table 5-1. Energy barriers and bond lengths for AM1 and B3LYP	107
Table 5-2. Setup of the system for QM/MM simulations.....	107
Table 5-3. Active atoms in the protein.....	108
Table 5-4. Hydrogen bond interactions for the QM/MM optimised reactants in Å ..	110

Table 5-5. Hydrogen bonds between the chloride and ambient water in the active site in the AM1/AMBER03 transition state structure	111
Table 5-6. Hydrogen bond distances in the QM/MM simulated products	112
Table 5-7. Reaction energies for the protonation of Lys78 with AM1/AMBER03 and TPSS/def2-TZVP+/AMBER03	114
Table 5-8. Reaction energies and barriers relative to the free reactants in kJ/mol of all the steps along the mechanism of chlorination of tryptophan in PrnA using QM/MM methods	126
Table 6-1. Experimental (aqueous solution) ^{135,136} and calculated (gas phase) spectra with M06-2X/6-311+G(2df, p) of pyridoxal 5'-phosphate	135
Table 6-2. Experimental ^{135,136} and calculated spectra using with M06-2X/6-311+G(2df, p) method of pyridoxal in IEFPCM continuum water	140
Table 6-3. Energies of the A and B relative to the most stable tautomers at the M06-2X/6-311G(2df, p) level in vacuum.....	142
Table 7-1. Selected X-ray data for fluoroimines 6 , 7 , and 8 , produced by Gilmour and co-workers ¹⁶⁹	158
Table 7-2. Calculated relative conformational energies and torsion angles of ethylimines derived from selected aldehydes RCOH. The results were obtained with M06-2X/6-311G(2df, p) in vacuum.....	160
Table 7-3. Calculated relative stabilities of tautomeric and protonation forms of neutral pyridoximine 8 in vacuum and water. The relative energies of the most stable conformer of each form at the M06-2X/6-311+G(2df, p) level in vacuum and continuum water respectively	164
Table 7-4. Calculated relative conformational energies of different forms of β -fluoropyridoximine 8 . The results were obtained with M06-2X/6-311G(2df, p) in vacuum and in continuum water (values in parentheses). Energies are relative to the most stable conformer of the respective form in vacuum or water, respectively	165

Acknowledgements

This thesis would not be completed without the help of any of the persons who I am about to acknowledge. First, I would like to express my gratitude to my supervisor Dr. Hans Martin Senn for all of his support and encouragement during the last four years. I am very grateful for all the time he spent helping me understand the theoretical chemistry. Secondly, I want to thank Prof. John McGrady and all the member of his group for creating pleasant working atmosphere while they were still in Glasgow. I would also like to acknowledge the project students working on the PrnA project (Christopher and Sarath), as well as all the other project students who shared the office with me. My thankfulness is extended to the members, students as well as staff, at the Glasgow University who have all made my stay in Glasgow a lot easier. I want to extend my gratitude to Craig and Derek for their help to correct my endless mistakes in English and make this manuscript easier to read. I promise, I will read, “Guide to the Queen’s English” one day.

Last but not least I want to thank my parents and my brother for encouraging me and being supportive in anything I have set my mind to in the last 29 years.

Author's Declaration

The work presented in this thesis is original, except if it was stated otherwise by the added references. No part of this work has previously been submitted for any other degree.

Some of the results presented here have appeared in the following publication:

Theoretical and X-ray Crystallographic Evidence of a Fluorine-Imine Gauche Effect: An Addendum to Dunathan's Stereoelectronic Hypothesis, Sparr, C.; Salamanova, E.; Schweizer, W. B.; Senn, H. M.; Gilmour, R. *Chem.-Eur. J.* **2011**, *17*, 8850.

Evdokiya K. Salamanova

To my father Krasimir...

Returning home, to childhood's only place,
when evening's timid flame is slowly fading
and silent night, in silent, soft embrace
unfolds to comfort mournful and forsaken.
To throw aside the weary load of hard
and blackened days, bequeathed by grief and care,
and stir with gentle footsteps through the yard
a gentle joy to see a loved one there.

Dimcho Debelyanov

Да се завърнеш в бащината къща,
когато вечерта смирено гасне
и тихи пазви тиха нощ разгръща
да приласкае скръбни и нещастни.
Кат бreme хвърлил черната умора,
що безутешни дни ти завещаха –
ти с плахи стъпки да събудиш в двора
пред гостенин очакван радост плаха.

Димчо Дебелянов

Chapter 1 Methods

1.1 Quantum Chemistry Theory

1.1.1 Electronic structure and wavefunction

Fundamental in quantum mechanics is to describe the electronic structure of many-body system. Central equation in this area is the time-independent, non-relativistic Schrödinger equation:²

$$\hat{H}\Psi = E\Psi \quad (1-1)$$

In this equation \hat{H} represents the Hamiltonian operator, Ψ is the wavefunction and E is the energy, eigenvalue of the Hamiltonian operator. The Hamiltonian can be written as a sum of kinetic and potential energies of the nuclei and electrons:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn} \quad (1-2)$$

where the first two operators correspond to the kinetic energies of the nuclei (\hat{T}_n) and electrons (\hat{T}_e). The next expressions represent the interactions between the nuclei and the electrons (\hat{V}_{ne}), the electron-electron Coulomb interaction (\hat{V}_{ee}) and the nucleus-nucleus repulsion term (\hat{V}_{nn}). The mathematical representation of the \hat{H} can be written as:

$$\hat{H} = \sum_{i=1}^N \frac{\hbar^2}{2m_e} \nabla_i^2 + \sum_{A=1}^M \frac{\hbar^2}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{e^2 Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{e^2}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{e^2 Z_A Z_B}{R_{AB}} \quad (1-3)$$

where A and B run over the M nuclei and i and j- over the N electrons. M_A is the mass of the nucleus, Z the atomic number, \hbar is the Planck's constant divided by 2π , and m_e is the mass of the electron. r_{iA} , r_{ij} and R_{AB} are respectively the distances between two particles. The Laplace operator ∇_i^2 denotes the sum of differential operators with respect to the Cartesian coordinates:

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \quad (1-4)$$

The wavefunction Ψ is a function of 3 (M+N) spatial coordinates and N spin coordinates. Due to this, the exact solution to the Schrödinger equation is known for the hydrogen atom, but is not known for any system bigger than this. In the cases where larger systems are being looked at, there is the need to find as accurate an approximation as possible, in order to obtain satisfactory results.

1.1.2 Born-Oppenheimer approximation

In the Born-Oppenheimer approximation³ it is assumed that the nuclei of the system are fixed in position and only the movement of the electrons needs to be accounted for. The mass of a proton is 1800 greater than the mass of the electron and therefore the nuclei move at a rate considerably slower than the rate at which the electrons move. The kinetic energy in this case is almost zero and the nucleus-nucleus repulsion is constant. As a result of this, the Hamiltonian is reduced to the electronic Hamiltonian:

$$\hat{H}_{elec} = \sum_{i=1}^N \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{e^2 Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{e^2}{r_{ij}} = \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee} \quad (1-5)$$

where \hat{V}_{nn} is now omitted because it is a constant. Using the electronic Hamiltonian to solve the Schrödinger equation will yield the electronic energy E_{elec} and the electronic wave function Ψ_{elec} :

$$\hat{H}_{elec} \Psi_{elec} = E_{elec} \Psi_{elec} \quad (1-6)$$

The nuclear energy is a constant:

$$V_{nn} = \sum_{A=1}^M \sum_{B>A}^M \frac{e^2 Z_A Z_B}{R_{AB}} \quad (1-7)$$

and the total energy of the system is:

$$E_{tot} = E_{elec} + V_{nn} \quad (1-8)$$

Without the Born-Oppenheimer approximation, the calculation of the total energy would be very computationally demanding.

1.1.3 The Slater determinant

All the electrons are characterised by a spin quantum number. Their spin could be $\pm 1/2$ and their spin functions (α and β) are orthonormal:

$$\langle \alpha | \alpha \rangle = 1 \quad \langle \beta | \beta \rangle = 1 \quad \langle \alpha | \beta \rangle = 0 \quad \langle \beta | \alpha \rangle = 0 \quad (1-9)$$

According to the *Pauli exclusion principle*, two electrons cannot have the same set of quantum numbers. The electrons are fermions and the interchange of any two electrons brings about a change in the sign of the wavefunction Ψ i.e. the function is antisymmetric:

$$\Psi(\vec{x}_1, \vec{x}_2) = -\Psi(\vec{x}_2, \vec{x}_1) \quad (1-10)$$

in the coordinates both, the spatial and the spin coordinates are encountered:

$$\vec{x}_i = (\vec{r}_i, s_i) \quad (1-11)$$

A way of writing the appropriate functional form of the ground state (Ψ_0) of a polyelectronic system with N number of electrons and $\chi_1, \chi_2, \dots, \chi_N$ spin orbitals, satisfying the antisymmetry principle, is the Slater determinant, ϕ_{SD} :⁴

$$\Psi_0 \approx \Phi_{SD} = \frac{1}{N!} \begin{vmatrix} \chi_1(\vec{x}_1) & \chi_2(\vec{x}_1) & \cdots & \chi_N(\vec{x}_1) \\ \chi_1(\vec{x}_2) & \chi_2(\vec{x}_2) & \cdots & \chi_N(\vec{x}_2) \\ \vdots & \vdots & & \vdots \\ \chi_1(\vec{x}_N) & \chi_2(\vec{x}_N) & & \chi_N(\vec{x}_N) \end{vmatrix} \quad (1-12)$$

Each of the one-electron spin orbitals $\chi_i(\vec{x}_i)$ can be written as a product of its spatial orbital $\phi_i(\vec{r})$ and one of the spin functions, either $\alpha(s)$ or $\beta(s)$:

$$\chi(\vec{x}) = \phi(\vec{r})\sigma(s) \quad \sigma = \alpha, \beta \quad (1-13)$$

1.1.4 Density Functional Theory (DFT)

1.1.4.1 Hohenberg-Kohn theorems

Using the Schrödinger equation and its Hartree-Fock approximation to solve chemical problems is rather difficult and computationally costly. Based on the non-observable wave function, they treat every electron individually. On the other hand, DFT created by Hohenberg and Kohn in 1964 uses the physical observable electron density (ρ) to describe the energy of a many-electron systems.⁵ The electron density reduces the problem to three dimensions and can be measured experimentally by X-ray diffraction. Integrating the electron density over all space gives the total number of electrons:

$$N = \int \rho(\vec{r}) d\vec{r} \quad (1-14)$$

Hohenberg and Kohn derived two theorems, which laid the foundations of the Density Functional Theory. The first theorem, called the “Existence” theorem, states that the external potential $V_{ext}(\vec{r})$ determines the ground state density. Hohenberg and Kohn

used *reductio ad absurdum* to prove their theorem i.e. the assumption of the opposite leads to a contradiction, which proves that it is true. We consider two different external potentials, V_{ext} and V'_{ext} , which describe the same non-degenerate ground state density ρ_0 . The two external potentials are part of the Hamiltonian \hat{H} , which gives the value of the ground state wave function Ψ_0 and its associated eigenvalue of energy E_0 . The different external potentials are associated with two different Hamiltonians \hat{H} and \hat{H}' , which correspond to two wave functions Ψ_0 and Ψ'_0 and their energies E_0 and E'_0 . The variational principle states that:

$$E_0 < \langle \Psi'_0 | \hat{H} | \Psi'_0 \rangle = \langle \Psi'_0 | \hat{H} | \Psi'_0 \rangle + \langle \Psi'_0 | \hat{H} - \hat{H}' | \Psi'_0 \rangle \quad (1-15)$$

this integral can be written in terms of the ground state density:

$$E_0 < E'_0 + \int [V_{ext}(\vec{r}) - V'_{ext}(\vec{r})] \rho_0(\vec{r}) d\vec{r} \quad (1-16)$$

If we utilise the equation above for V'_{ext} we get the corresponding equation:

$$E'_0 < E_0 + \int [V'_{ext}(\vec{r}) - V_{ext}(\vec{r})] \rho_0(\vec{r}) d\vec{r} \quad (1-17)$$

Adding the two equations will lead us to the conclusion:

$$E_0 + E'_0 < E'_0 + E_0 \quad (1-18)$$

which by itself is a contradiction and proves that it is not possible that two different external potentials V_{ext} may yield the same ground state electron density ρ_0 . Via this

proof, it can be categorically stated that the external potential is a unique functional of the electron density.

The ground energy expression includes two components- those, which depend on the actual system, such as the nuclei-electron attraction potential energy $V_{Ne}[\rho_0]$, and those, which are system independent - the electron kinetic energy $T[\rho_0]$ and the electron-electron repulsion potential energy $E_{ee}[\rho_0]$:

$$E_0[\rho] = \int \rho_0(\vec{r}) V_{Ne} d\vec{r} + T[\rho_0] + E_{ee}[\rho_0] \quad (1-19)$$

The system independent part of the equation can be collected together and placed into a functional called Hohenberg-Kohn functional $F_{HK}[\rho_0]$:

$$E_0[\rho] = \int \rho_0(\vec{r}) V_{Ne} d\vec{r} + F_{HK}[\rho_0] \quad (1-20)$$

The electron-electron interaction functional $E_{ee}[\rho]$ can be written as:

$$E_{ee}[\rho] = J[\rho] + E_{ncl}[\rho] \quad (1-21)$$

where $J[\rho]$ is the classical Coulomb integral and $E_{ncl}[\rho]$ is the non-classical representation of the electron self-interaction, exchange and Coulomb correlation.

The second Hohenberg-Kohn theorem represents the variational principle. This theorem states that $F_{HK}[\rho]$ gives the lowest energy only if the input density is the true ground state density, ρ_0 :

$$E_0 \leq E[\tilde{\rho}] = T[\tilde{\rho}] + E_{Ne}[\tilde{\rho}] + E_{ee}[\tilde{\rho}] \quad (1-22)$$

This states that any trial density $\tilde{\rho}(\vec{r})$, which is associated with some external potential \tilde{V}_{ext} , gives energy greater than or equal to the true ground state energy E_0 .

1.1.4.2 Kohn-Sham approach

Year later, after the Hohenberg-Kohn theorems appeared, in 1965 Kohn and Sham introduced an approach for describing the kinetic energy within the Density Functional Theory in a more accurate and efficient way.⁶ They presented the idea of a non-interacting system built from one-electron functions (orbitals) similar to the concept of the Slater determinant Φ_{SD} in the Hartree-Fock method.

The Hamiltonian of the non-interacting reference system in KS approach takes the following expression:

$$\hat{H}_s = -\frac{1}{2} \sum_i^N \nabla_i^2 + \sum_i^N V_s(\vec{r}_i) \quad (1-23)$$

where the electrons in the non-interacting system are in the field of an effective local potential $V_s(\vec{r})$.

Analogously to the Hartree-Fock the spin orbitals can be presented as follows:

$$\hat{f}^{KS} \varphi_i = \varepsilon_i \varphi_i \quad (1-24)$$

and the one-electron Kohn-Sham operator is defined as:

$$\hat{f}^{KS} = -\frac{1}{2} \nabla^2 + V_s(\vec{r}) \quad (1-25)$$

Connecting our non-interacting reference system, built from KS orbitals, to the real system with interacting electrons is possible by choosing an effective local potential $V_s(\vec{r})$ such as the sum of the moduli of the KS squared orbitals, which will give the exact ground state density of the real system:

$$\rho_s(\vec{r}) = \sum_i^N \sum_s |\phi_i(\vec{r}, s)|^2 = \rho_0(\vec{r}) \quad (1-26)$$

Going back to the original idea of using the kinetic energy on a non-interacting system to express the total energy, Kohn and Sham introduced the following separation of the functional to encounter the difference between the true kinetic energy and the kinetic energy of the non-interacting system:

$$F[\rho(\vec{r})] = T_s[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{xc}[\rho(\vec{r})] \quad (1-27)$$

where E_{xc} is the exchange-correlation functional and it is determined as:

$$E_{xc}[\rho] \equiv (T[\rho] - T_s[\rho]) + (E_{ee}[\rho] - J[\rho]) = T_c[\rho] + E_{ncl}[\rho] \quad (1-28)$$

where the kinetic energy of the non-interacting system T_s is added to the non-classical electrostatic interactions and it is not part of the true kinetic energy T_c . Thus the exchange-correlation functional contains everything that is unknown in the system.

1.1.4.3 Exchange-correlation functionals

1.1.4.3.1 The Local Density Approximation (LDA)

These types of functional are based on the so-called uniform electron gas in which the electron density is a constant everywhere. The total exchange-correlation energy can then be integrated all over the space:⁷

$$E_{xc}[\rho] = \int \rho(\vec{r}) \epsilon_{xc}(\rho(\vec{r})) d\vec{r} \quad (1-29)$$

in the equation $\epsilon_{xc}(\rho(\vec{r}))$ is exchange-correlation energy per electron as a function of the density in the uniform gas.

The exchange-correlation energy per electron of the uniform electron gas is known from different approaches such as quantum Monte Carlo methods. An example of a LDA functional is VWN (Vosko, Wilk and Nuisair).⁸

1.1.4.3.2 The Generalised Gradient Approximation (GGA)

The Local Density Approximation is insufficiently accurate to solve the majority of the problems in chemistry. That was the reason for developing a new type of functionals, the so-called “non-local” functionals. Commonly used types of non-local functionals are the gradient-corrected functionals (GGA), which depend not only on the value of the density at each point of the space but on its gradient as well. The correlation and exchange are usually separated in these functionals. Usually they are built by adding a term to a LDA functional i.e.,

$$E_X^{GGA}[\rho(\vec{r})] = E_X^{LDA} \sum_{\sigma} \int F(s_{\sigma}) \rho_{\sigma}^{4/3}(\vec{r}) d\vec{r} \quad (1-30)$$

That expression for the exchanged part of E_{xc} is based on the reduced gradient F with spin σ . s_{σ} inhomogeneity parameter equal to:

$$s_{\sigma}(\vec{r}) = \frac{|\nabla \rho_{\sigma}(\vec{r})|}{\rho_{\sigma}^{4/3}(\vec{r})} \quad (1-31)$$

Widely used GGA functionals are the Lee-Yang-Parr (BLYP),⁹ the Filatov and Thiel (FT97),¹⁰ the Perdew-Wang PW91,¹¹ Perdew-Burke-Ernzerhof functional,¹² etc.

1.1.4.3.3 Hybrid functionals

The important feature of DFT compared to Hartree-Fock is the inclusion of the correlation effects. However, the Hartree-Fock method provides an exact treatment of the exchange contribution. The exchange contribution to the E_{XC} energy is significantly larger than the correlation contribution. Thus using the Slater determinant of KS type of orbitals, it is possible to give more accurate description of the exchange-correlation. The E_{XC} energy in this case has an exact exchange E_X^{exact} and approximate correlation E_C^{KS} terms:

$$E_{XC} = E_X^{exact} + E_C^{KS} \quad (1-32)$$

This “half-half” theory provides better results than the alternative methods. Becke derived the mixed exchange-correlation HF, LSDA (Local Spin Density Approximation) and GGA functionals:

$$E_{XC} = E_{XC}^{LSDA} + a_0(E_X^{exact} - E_X^{LSDA}) + a_x \Delta E_X^{GC} + a_c \Delta E_C^{GC} \quad (1-33)$$

where E_x^{LSDA} is the exchange energy under local spin density approximation, ΔE_x^{GC} is the gradient correction for exchange, ΔE_c^{GC} is the gradient correction for the correlation and a_0 , a_x and a_c are coefficients, obtained from least-square fitting to experimental data.

Currently the most popular hybrid functionals are B3LYP⁹, M05(6)-X(2X),¹³ the TPSS hybrid functional-TPSSH.¹⁴

1.1.5 Time-Dependent Density Functional Theory (TDDFT)

The foundations of TDDFT were laid by Runge and Gross in 1984,^{15,16} who derived a Hohenberg-Kohn-like theorem for the time-dependent Schrödinger equation:

$$i \frac{\partial}{\partial t} \Psi(\vec{r}, t) = \hat{H}(\vec{r}, t) \Psi(\vec{r}, t) \quad (1-34)$$

The main variable in TDDFT is as in DFT - not the wavefunction but the electron density. The Runge-Gross theorem proves that there is a one-to-one mapping between the external time-dependent potential V_{ext} of an electric field and the time-dependent density. If the external time-dependent potential is “small”, the time-dependent Kohn-Sham equations’ solutions can be avoided by using the linear response theory. This is the case for photoabsorption spectra. In the case where there is a “strong” external field (laser beam), the response is no longer linear and a full solution to the time-dependent Kohn-Sham equations has to be sought.

Runge and Gross applied the KS approach to a time-dependent Schrödinger equation:

$$i \frac{\partial}{\partial t} \varphi_i(\vec{r}, t) = \hat{H}_{KS}(\vec{r}, t) \varphi_i(\vec{r}, t) \quad (1-35)$$

where φ_i represents the KS orbitals and \hat{H}_{KS} is the KS Hamiltonian, which can be written as:

$$\hat{H}_{KS}(\vec{r}, t) = -\frac{\nabla^2}{2} + V_{KS}[n](\vec{r}, t) \quad (1-36)$$

where V_{KS} is the KS potential, that as in the KS scheme for the ground state can be written as a sum of three terms, the external potential, the classical electrostatical interaction and the exchange-correlation term :

$$V_{KS}[n](\vec{r}, t) = V_{ext}(\vec{r}, t) + V_{Hartree}[n](\vec{r}, t) + V_{XC}[n](\vec{r}, t) \quad (1-37)$$

The excitation energies of a system can be obtained by knowing how the system responds to a small time-dependent perturbation. At time $t < t_0$ the time-dependent potential is zero i.e. the system is subject only to the nuclear external potential and the density is the ground state density $n^{(0)}$. At t_0 we apply an external field with $V^{(1)}(\vec{r}, t)$ and the total external potential in this case is $V(\vec{r}, t) = V^{(0)}(\vec{r}) + V^{(1)}(\vec{r}, t)$. The external potential will cause a change in the density. Then we can present the density as an expansion of perturbative series:

$$n(\vec{r}, t) = n^{(0)}(\vec{r}) + n^{(1)}(\vec{r}, t) + n^{(2)}(\vec{r}, t) + \dots, \quad (1-38)$$

where $n^{(1)}$ is a component of $n(\vec{r}, t)$ that depends linearly on $V^{(1)}$, $n^{(2)}$ depends quadratically, etc. If the perturbation is weak, we can only be concerned with the linear term.

The response of the density to a weak optical external field in TDDFT is characterised by the imaginary polarizability, part of the photo-absorption cross-section tensor:

$$\sigma(\omega) = \frac{4\pi\omega}{c} \text{Im} \sum_{\gamma} \alpha_{\gamma}(\omega) \quad (1-39)$$

where ω stands for the angular frequency, $\sigma(\omega)$ is the photoabsorption cross section, c is the speed of light $\gamma = x, y, z$ and the polarisability $\alpha_{\gamma}(\omega)$ is given by the following equation; σ, σ' are the spin variables:

$$\alpha_{\gamma}(\omega) = - \int d^3\vec{r} \int d^3\vec{r}' r_{\gamma} \sum_{\sigma\sigma'} \chi_{\sigma\sigma'}(\vec{r}, \vec{r}', \omega) r'_{\gamma} \quad (1-40)$$

where χ is the linear response function, which measures the change in the density δn perturbed by a change in the external field:

$$\delta n_{\sigma}(\vec{r}, \omega) = \sum_{\sigma'} \int d^3\vec{r}' \chi_{\sigma\sigma'}(\vec{r}, \vec{r}', \omega) \delta V_{ext\sigma'}(\vec{r}', \omega) \quad (1-41)$$

in terms of Kohn-Sham orbitals the linear response function can be written as follows:

$$\chi_{KS\sigma\sigma'}(\vec{r}, \vec{r}', \omega) = \delta_{\sigma\sigma'} \sum_{jk} (f_{k\sigma} - f_{j\sigma}) \frac{\varphi_{j\sigma}(\vec{r}) \varphi_{j\sigma}^*(\vec{r}') \varphi_{k\sigma}(\vec{r}') \varphi_{k\sigma}^*(\vec{r})}{\omega - (\varepsilon_{j\sigma} - \varepsilon_{k\sigma}) + i\eta} \quad (1-42)$$

$\varphi_{j\sigma}$ and $\varepsilon_{j\sigma}$ are the ground-state Kohn-Sham orbitals and energies, f_j is the occupation number of the j th orbital, and η is a positive infinitesimal.

Using the KS potential it is possible to obtain the linear change in the KS potential:

$$\begin{aligned} \delta V_{KS}(\vec{r}, \omega) = & \delta V_{ext\sigma}(\vec{r}, \omega) + \int d^3r' \frac{\delta n(\vec{r}', \omega)}{|\vec{r} - \vec{r}'|} + \\ & + \sum_{\sigma'} \int d^3r' f_{XC\sigma\sigma'}(\vec{r}, \vec{r}', \omega) \delta n_{\sigma'}(\vec{r}', \omega) \end{aligned} \quad (1-43)$$

In the previous equation the exchange-correlation kernel f_{XC} was introduced and the using Fourier transform the XC kernel can be presented as follows:

$$f_{XC\sigma\sigma'}[n](\vec{r}\vec{r}', t t') = \frac{\delta V_{XC}(\vec{r}, t)}{\delta n(\vec{r}', t')} \quad (1-44)$$

The density response function (I-41) has poles at frequencies that correspond to the energies of the interacting system and the KS response function has poles at the difference of KS eigenvalues ($E_{j\sigma} - E_{k\sigma}$). Thus the change in the density can be expressed as KS transitions:

$$\delta n(\vec{r}, \omega) = \sum_{ia} \left[\xi_{ia\sigma}(\omega) \varphi_{a\sigma}^*(\vec{r}) \varphi_{i\sigma}(\vec{r}) + \xi_{ai\sigma}(\omega) \varphi_{a\sigma}(\vec{r}) \varphi_{i\sigma}^*(\vec{r}) \right] \quad (1-45)$$

where i stands for the occupied. a for the virtual state and ξ are the solution vectors ($\xi_{ia\sigma}$ is the excitation particle- hole and $\xi_{ai\sigma}$ is the deexcitation hole- particle).

Inserting this expression in equation (I-41) yields two coupled matrix equations:

$$\left[\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix} - \omega \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \right] \begin{pmatrix} X \\ Y \end{pmatrix} = - \begin{pmatrix} \delta v \\ \delta v^* \end{pmatrix} \quad (1-46)$$

where the matrices A and B are defined in the product space of the occupied and virtual states and can be defined as :

$$\begin{aligned} A_{ia\sigma, i'a'\sigma'} &= \delta_{ii'} \delta_{aa'} \delta_{\sigma\sigma'} (\epsilon_{a\sigma} - \epsilon_{i\sigma}) + K_{ia\sigma, i'a'\sigma'}, \\ B &= K_{ia\sigma, i'a'\sigma'}, X = \xi_{ia\sigma}, Y = \xi_{ai\sigma} \end{aligned} \quad (1-47)$$

and

$$K_{ia\sigma, i'a'\sigma'} = \int d^3\vec{r} \int d^3\vec{r}' \varphi_{a\sigma}^*(\vec{r}) \varphi_{i\sigma}(\vec{r}) \left[\frac{1}{|\vec{r} - \vec{r}'|} + f_{XC\sigma\sigma'}(\vec{r}, \vec{r}', \omega) \right] \varphi_{a'\sigma'}(\vec{r}) \varphi_{i'\sigma'}^*(\vec{r}) \quad (1-48)$$

with

$$\delta v_{ai\sigma}(\omega) = \int d\vec{r} \varphi_{ai\sigma}(\vec{r}) \varphi_{ai\sigma}^*(\vec{r}) \delta V_{ext}(\vec{r}, \omega) \quad (1-49)$$

The eigenvalues of these equations are the squared of the excitation energies and eigenvectors can be used to calculate the oscillator strength.

Most of TDDFT calculations agree with the experiments. No physical theory works for all the systems and TDDFT is not an exception. TDDFT is an exact reformulation of the time-dependent Schrödinger equation and in that sense it fails where the QM does not work. The key approximation is the approximation of V_{XC} and therefore the errors in the calculations depend on the functional that was used.

1.1.6 Semi-empirical Methods

Back in the early days of theoretical and computational chemistry using Hartree-Fock formalism was a challenging task, due to resource limitations. The first step in the simplification of the problem was to consider only the valence electrons of the system and the core electrons are treated by reducing the nuclear charge or by introducing functions to model the combined repulsion between the nuclei and the core electrons. This type of approximation is part of the Extended Hückel Theory (EHT).¹⁷ The majority of the semi-empirical methods use just s- and p-basis function and the functions are represented by Slater-type orbitals (STO-orbitals). The second so-called numerical approximation is made in order to set some integrals in the Hartree-Fock

calculation to zero. The most computationally demanding part of the Hartree-Fock calculations is the solution to the two-electron (four-index) integrals J and K, responsible for the electron-electron repulsion and exchange interaction. These integrals scale the calculations to the fourth power of the number of basis functions N^4 if N is the number of basis functions. With respect to the Coulomb and exchange integrals, which describe the interaction between two electrons in the space defined by their basis functions, when the two electrons are far away from each other, the basis functions in these instances can be set to zero. The same could be done with one-electron integrals describing the interaction between the nuclei and electrons but that would not speed up the calculation as much as the cancelation of the J and K integrals does. Despite this, the semi-empirical methods were still designed because of the lack of computational power available. They are still widely-used in the present day due to their speed and efficiency.¹⁸

1.1.7 NDDO formalism

NDDO (*neglect of diatomic differential overlap*) formalism neglects the differential overlap between atomic orbitals on different atoms. The main idea of this method is a replacement representation of the elements in the HF secular equation. The secular equation takes the following form of a determinant:

$$\begin{vmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \cdots & H_{1N} - ES_{1N} \\ H_{21} - ES_{21} & H_{22} - ES_{22} & \cdots & H_{2N} - ES_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1} - ES_{N1} & H_{N2} - ES_{N2} & \cdots & H_{NN} - ES_{NN} \end{vmatrix} = 0 \quad (1-50)$$

where $H_{\mu\nu}$ and $S_{\mu\nu}$ are the “resonance” integral and the “overlap” integral, respectively. These integrals can be expressed as follows:

$$H_{\mu\nu} = \int \varphi_{\mu} \hat{H} \varphi_{\nu} d\vec{r} \quad (1-51)$$

$$S_{\mu\nu} = \int \varphi_{\mu} \varphi_{\nu} d\vec{r} \quad (1-52)$$

where φ_i and φ_j are the atomic basis set functions and $a_{\mu\nu}$ is the basis function coefficient in the linear combination of the atomic orbitals (LCAO). In the Roothan-Hall equations these integrals take the form of matrices and they are called Fock and overlap matrices ($F_{\mu\nu}$ and $S_{\mu\nu}$). The overlap between the basis functions is not necessarily zero, e.g. they can be situated on different atoms. The energy of the system is then:

$$E = \frac{\sum_{\mu\nu} a_{\mu\nu} H_{\mu\nu}}{\sum_{\mu\nu} a_{\mu\nu} S_{\mu\nu}} \quad (1-53)$$

The purpose of the secular equation is to find the coefficients in the equation:

$$\sum_{\mu=1}^N a_{\mu\nu} (H_{\mu\nu} - ES_{\mu\nu}) = 0 \quad (1-54)$$

To find the unknown, in this case the coefficients, we form the secular determinant and set it to zero.

In the NDDO formalism the Fock matrix elements are equal to:

$$F_{\mu\mu} = H_{\mu\mu}^{core} + \sum_{\lambda \text{ on } A} \sum_{\sigma \text{ on } A} \left[P_{\lambda\sigma}(\mu\mu | \lambda\sigma) - \frac{1}{2} P_{\lambda\sigma}(\mu\lambda | \mu\sigma) \right] + \sum_{B \neq A} \sum_{\lambda \text{ on } B} \sum_{\sigma \text{ on } B} P_{\lambda\sigma}(\mu\mu | \lambda\sigma) \quad (1-55)$$

For μ and ν on the same atom (both on A):

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda \text{ on } A} \sum_{\sigma \text{ on } A} \left[P_{\lambda\sigma}(\mu\nu | \lambda\sigma) - \frac{1}{2} P_{\lambda\sigma}(\mu\lambda | \nu\sigma) \right] + \sum_{B \neq A} \sum_{\lambda \text{ on } B} \sum_{\sigma \text{ on } B} P_{\lambda\sigma}(\mu\nu | \lambda\sigma) \quad (1-56)$$

For μ and ν on different atoms (A and B):

$$F_{\mu\nu} = H_{\mu\nu}^{core} - \frac{1}{2} \sum_{\lambda \text{ on } B} \sum_{\sigma \text{ on } A} P_{\lambda\sigma}(\mu\sigma | \nu\lambda) \quad (1-57)$$

where μ and ν are the basis functions for the first electron and λ and σ are the basis functions for the other, P is the density matrix formed from the basis functions coefficients, H^{core} is the core Hamiltonian that includes the one-electron integrals, which do not depend on P . The kinetic energy with the removal of the interactions of nuclei other than A gives an empirically defined parameter $U_{\mu\nu}$.

The two most widely used NDDO models are MNDO¹⁹ and AM1.²⁰

1.1.7.1.1 AM1

The *Austin Model* was designed by Dewar and co-workers in 1985 to eliminate some of the errors associated with MNDO in the calculations of the electron-electron repulsion. In a distance of the sum of the van der Waals radii (2-3 Å) between two atoms, the repulsion was too high. The solution was to modify the core-core functions, using Gaussian repulsive and attractive functions. The attractive functions were designed to be larger in the region where the repulsive interactions were too great. The repulsive functions were centred at smaller internuclear distances. After the modifications the core-core term took the following expression:

$$E_{AB} = E_{MNDO} + \frac{Z_A Z_B}{R_{AB}} \times \\ \times \left\{ \sum_i K_{Ai} \exp \left[-L_{Ai} (R_{AB} - M_{Ai})^2 \right] + \sum_j K_{Bj} \exp \left[-L_{Bj} (R_{AB} - M_{Bj})^2 \right] \right\} \quad (1-58)$$

The additional terms are spherical Gaussian functions, determined by the L parameter. The original MNDO parameters were improved after optimisations.

1.2 Molecular Mechanics (Empirical force fields)

Many of the systems encountered in computational chemistry are too large to be considered, using quantum chemistry methods. In this case an empirical force field can be applied. Molecular mechanics is the preferred method for many-particle systems as proteins, nucleic acids, polymers, etc., when the calculations are considerably time-consuming. The reason for the fast speed and simplicity of the calculation is the fact that the electronic structure is ignored. The energy of the systems is calculated as a function of the nuclear position. The atoms are represented by balls and the bonds by springs. Molecular mechanics cannot describe properties, which are related to the electronic structure of the molecule, such as bond-forming or bond-breaking, electronic spectra, etc. It gives satisfactory results for systems with many atoms and for conformational analysis. The parameters in the force fields are fitted to experimental data or high-level *ab initio* calculations. The main principle in the force fields is the transferability, e.g. the parameters are developed for small type of molecules and then are applied to bigger structures.^{7,18}

The potential energy of the force field can be represented as a five-component picture:

$$E_{MM} = \underbrace{E_{bonds} + E_{angles} + E_{torsions}}_{bonded} + \underbrace{E_{vdw} + E_{electr}}_{non-bonded} \quad (1-59)$$

The first three terms correspond to the bonded terms and consist of bond, angle and torsion terms. The last two terms are the non-bonded components, known as the van der Waals and electrostatic interactions.

The bond energy obeys Hooke's law and its potential energy curve has a parabolic shape. In Hooke's formula, the potential energy varies with the square of the

displacement l from the reference bond length l_0 and the strength of the bond is expressed by the force constant k :

$$E_{bonds} = \sum_{bonds} \frac{k}{2} (l - l_0)^2 \quad (1-60)$$

The true bond potential is not harmonic, because the energy does not have an infinite positive value at infinite distance. A better representation is the Morse potential:

$$E_{bonds} = \sum_{bonds} D_e \{1 - \exp[-a(l - l_0)]\}^2 \quad (1-61)$$

D_e is the depth of the minimum and $a = \omega \sqrt{\mu / 2D_e}$ where μ is the reduced mass and ω is the frequency of the bond vibration. ω is related to the force constant by $\omega = \sqrt{k / \mu}$. The Morse potential requires three parameters per bond, which limits its computational efficiency.

Hooke's law also describes the angle potential with the force constant k and the angle reference value θ_0 :

$$E_{angles} = \sum_{angles} \frac{k}{2} (\theta - \theta_0)^2 \quad (1-62)$$

Bond stretching and angle bending require a lot of energy for their distortion. The main changes to the bonded terms during an energy minimisation are due to a change in the torsion angles. Not all the force fields have torsion potentials and the 1,4-atom interactions can be described by non-bonded interactions. However, most of the “organic” force fields have a torsion term with a cosine series expansion:

$$E_{\text{torsions}} = \sum_{\text{torsions}} \frac{k}{2} [1 + \cos(n\phi - \delta)] \quad (1-63)$$

where k is the barrier, n is the periodicity, ϕ is the torsion angle and δ is the phase. The periodicity values mark the number of minima on the rotation energy profile of 360° . The δ factor shows where these minimums are and their shift from zero degrees.

The van der Waals interactions are a balance between attractive and repulsive forces. The attractive forces are known as dispersive or London forces. The dispersive interactions are due to instantaneous induced dipoles. One such instantaneous dipole induces a dipole in a neighbouring part of the molecule and that is followed by an attractive interaction. The repulsive contribution to the van der Waals interaction occurs at small distances, around 3\AA (the sum of the van der Waals radii of the atoms). When the atoms approach one another at such small distances, the energy increases sharply. The explanation for the repulsive forces is in the Pauli exclusion principle, which states that two electrons cannot have the same four quantum numbers. The most commonly used function for van der Waals attractive and repulsive interactions is the Lennard-Jones 12-6 potential:

$$E_{\text{vdw}} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (1-64)$$

The Lennard-Jones parameters are the collision diameter σ (the distance at which the energy is zero) and well depth ϵ . The well depth gives the value of the energy at the equilibrium distance. Each atom can be assigned a partial charge q and the electrostatic interaction between two atoms A and B can be expressed, using Coulomb's law:

$$E_{electr} = \frac{q_A q_B}{\epsilon r_{AB}} \quad (1-65)$$

The choice of which force field to use depends on the system under investigation. There are specialised force fields for biomolecules (CHARMM, GROMOS, AMBER, OPLS), for organic molecules (MM2, MM3) and general force fields (UFF, ESSF). More specialised force fields are used most often, due to their accuracy.

1.3 QM/MM methods

The combined quantum mechanical/molecular mechanical methods make it possible to deal with systems of 1 000 000 atoms and it is also possible to look at reactions of bond-breaking, bond-forming, charge transfer or electronic excitations. Most of the systems treated by QM/MM are biopolymers, such as enzymes with an active centre, where a specific reaction takes place.²¹ The QM method treats the active site and its neighbours in the enzyme, forming the inner subsystem (I). The rest of the system, called the outer region (O) is described by force fields (MM). The inner and outer regions, combined, form the entire system, marked as (S). These two regions are also known as QM and MM regions, or primary and secondary subsystems. The expression for the QM/MM energy is not just a sum of the energies of the inner and outer regions, coupling terms are also needed. There are two schemes that can express the QM/MM energy- subtractive and additive. The subtractive scheme²² requires a MM calculation to be performed on the entire system, a QM calculation to be performed on the inner system and a MM calculation to be performed on the inner subsystem. The subtractive expression can be written as follows:

$$E_{QM/MM}^{sub}(S) = E_{MM}(S) + E_{QM}(I + L) - E_{MM}(I + L) \quad (1-66)$$

The subscript in the equation indicates the types of methods to be used and the letters in brackets represent the subsystem. The letter L marks the link atom. It caps the QM subsystem when it is cut through a covalent bond. The main idea is that the inner region is cut out and treated at the QM level. No explicit coupling terms are needed because it is handled at the MM level, but MM parameters for the inner region have to be derived, which can sometimes result in complications.

The additive scheme is the other way in which the system's energy could be expressed:

$$E_{Qm/MM}^{add}(S) = E_{MM}(O) + E_{QM}(I + L) + E_{QM-MM}(I, L) \quad (1-67)$$

In this case only the outer subsystem is treated at MM level, with $E_{QM-MM}(I, L)$ being the coupling term. The coupling describes both the bonded and non-bonded interactions occurring between the two subsystems. The most important of these interactions is the electrostatic interactions; it divides the QM/MM methods into two groups; the group of mechanical embedding and the group of electrostatic embedding.²³ In the mechanical embedding, the QM calculations are performed in the absence of the MM subsystem and the Coulomb interactions are calculated as MM electrostatics. The first drawback of the mechanical coupling is that when the electronic structure in the inner region changes, that does not mirror a charge distribution change of the MM parameters. The second drawback is connected with the lack of potential perturbation of the inner system by the charges of the outer region.

These problems are solved with the implementation of electrostatic embedding. The MM point charges are represented as a one-electron term in the Hamiltonian:

$$E_{QM-MM}^{el} = \sum_i^N \sum_{J \in O}^L \frac{q_J}{|r_i - R_J|} + \sum_{\alpha \in I+L}^M \sum_{J \in O}^L \frac{q_J Q_\alpha}{|R_\alpha - R_J|} \quad (1-68)$$

q_j represents the MM point charges at position R_j ; Q_α represent the QM nuclear charges at positions R_α ; i runs over N electrons, J over L point charges and α over M QM nuclei. In the electrostatic embedding, no MM electrostatic parameters are required. The electronic structure of the QM region is polarised by the environment. The disadvantages associated with electrostatic embedding are: 1) the computational cost is greater in comparison to the cost of mechanical coupling; 2) The MM charges are not polarised by the inner region electronic structure, they remain fixed and they do not change dynamically during the reaction; 3) The MM charges on the boundary can cause an overpolarisation of QM density.

The other interactions between the two regions include the van der Waals and bonded interactions, which are handled at the MM level. They are more easily described. They are either short range or occur between bonded atoms. The question is this: how these interactions are affected by the varying of the atom types present along the pathway in the QM region. A solution to this problem is to move the QM/MM boundary further away from the QM region.

An interesting problem is how to treat the boundary between the QM and the MM regions. In the ideal scheme, the boundary does not pass through a covalent bond e.g. a molecule solvated in explicit water, or biomolecular systems with substrates and co-factors, which are not covalently bound to the enzyme, catalysing the reaction. On many occasions, the frontier has to pass through a covalent bond and in some of the cases, even through an ionic bond. This causes problems as 1) In the QM region there is no saturated bond; 2) In the electrostatic embedding the QM region could be overpolarised by the MM point charges near the boundary; 3) Double counting of the interaction has to be avoided because the bonded MM terms include atoms from both regions.

There are several ways of treating the QM/MM boundary, but the most popular are the link-atom²⁴ and the frozen localised orbitals methods.²⁵ The link-atom method introduces an additional one-free-valence atom or group. Usually the link atom is a hydrogen atom. The boundary atom, in principle, should bring three new degrees of freedom but this is eliminated, placing it at a defined distance between the QM atom and MM atom at the frontier. This distance is precisely scaled by a factor, proposed by Morokuma and co-workers. In the link-atom scheme, the QM density in the

boundary could be overpolarised by the nearby MM point charges. Several solutions to this problem are known: deleting one-electron integrals associated with the link atom, omitting the MM charges at the boundary, or shifting the charge to neighbouring atoms. The deletion of the MM charges right next to the boundary is widely used in force fields like CHARMM22,²⁶ where the parameters are set to zero whole groups of atoms. Deleting the group of charges, connected to the MM boundary atom does not change the total charge of the system. For a force field without a neutral-group feature, a shift scheme can be applied. In such a force field the MM point charge is shifted onto the next atoms.

The frozen localised orbitals approaches place the hybrid orbitals on the QM or MM frontier atoms and keep them frozen so that they do not participate in the SCF. In the local self-consistent field (LSCF) approach,²⁷ the bond orbital consists of only the frontier atoms, it does not take part in the SCF and it is orientated along the QM-MM vector. Another method is the generalised hybrid orbital (GHO) method.²⁸ In essence, this involves an sp^3 hybrid orbital assigned to each of the MM boundary atoms. The orbital pointing towards the QM atom participates in the SCF.

1.4 Molecular dynamics

It is found that at 0 K, molecules vibrate and this is known as the zero-point vibrational effect. For small molecules, the potential energy surface (PES) is narrow and deep, so vibrating the molecule samples structures very close to the minimum. Large molecules such as proteins or nucleic acids are very flexible with many degrees of freedom. They have a wide variety of conformers, existing in physiological conditions. This range of structures will give different local minima on the PES at non-zero absolute temperature. To describe the motion of the molecules, in molecular dynamics Newton's second law of motion is used e.i. classical mechanics description.^{7,18,29}

$$F_i = m_i a_i \text{ (1-69)}$$

where F is the acting force, m is the mass, a is the acceleration and i runs over all the coordinates of all the particles in the system. The acceleration is the first derivative of the velocity and second derivative of the positions (R) with respect to the time:

$$a_i = \frac{dv}{dt} = \frac{d^2 R_i}{dt^2} = \ddot{R}_i \quad (1-70)$$

The updated coordinates are used to evaluate the potential energy. The coordinates as a function of time represent the trajectory of the system.

In dynamics, we propagate the present position $R(t)$ and velocities $V(t)$, during a time step Δt to obtain the new position $R(t + \Delta t)$ and velocities $V(t + \Delta t)$. The choice of the time step is very important as it depends on the fastest motion in the system. The time step has to be shorter in time than this motion, so that no change to the system is missed. Usually the fastest motion is C-H vibration with frequency $\nu=3000 \text{ cm}^{-1}$ with a period $T \approx 10 \text{ fs}$. This suggests that the time step should be 1-2 fs.

To try to obtain the new positions as a function of time along the trajectory, our first guess would be to use a second order Taylor expansion of the position coordinate:

$$R_i(t + \Delta t) = R_i(t) + \dot{R}_i(t)\Delta t + \ddot{R}_i(t)\frac{\Delta t^2}{2} + O(\Delta t^3) \quad (1-71)$$

But using this truncated Taylor expansion leads to an accumulation of errors and this brings the necessity of a new integrator. A “good” integrator would be an integrator that conserves the total energy through the simulation and is time reversible.

The most widely used method for integrating the equation of motion in molecular dynamics is the Verlet algorithm.³⁰ It is time-reversible and energy conserving, so we can trace back the positions along the trajectory and it keeps the total energy constant. The Verlet algorithm uses the positions and accelerations at time t , and the positions

from the previous time step $R(t - \Delta t)$, to calculate the new positions at time $t + \Delta t$, $R(t + \Delta t)$. We can write the two equations of the Taylor expansions in the forward and backward direction:

$$R_i(t + \Delta t) = R_i(t) + \dot{R}_i(t)\Delta t + \ddot{R}_i(t)\frac{\Delta t^2}{2} + \dddot{R}_i(t)\frac{\Delta t^3}{6} + O(\Delta t^4) \quad (1-72)$$

$$R_i(t - \Delta t) = R_i(t) - \dot{R}_i(t)\Delta t + \ddot{R}_i(t)\frac{\Delta t^2}{2} - \dddot{R}_i(t)\frac{\Delta t^3}{6} + O(\Delta t^4) \quad (1-73)$$

After adding the two equations we will obtain the Verlet algorithm:

$$R_i(t + \Delta t) = 2R_i(t) - R_i(t - \Delta t) + \ddot{R}_i(t)\Delta t^2 + O(\Delta t^4) \quad (1-74)$$

The velocities do not appear explicitly in the Verlet algorithm. As they are not used to propagate the positions, they are not available. The velocities are important because they are needed for the calculation of the kinetic energy, which is part of the total energy of the system. There are many ways to obtain them, but the most popular method is to divide displacement from previous to next steps by $2\Delta t$:

$$\dot{R}_i = \frac{R_i(t + \Delta t) - R_i(t - \Delta t)}{2\Delta t} \quad (1-75)$$

Another problem of the Verlet algorithm is that we have to subtract two almost equal numbers $2R(t)$ and $R(t-\Delta t)$.

A variation of the Verlet algorithm is the leap-frog method.³¹ First the velocities are calculated at half step and then the positions are obtained from these velocities by adding them to the positions at time t :

$$\dot{R}_i(t + \frac{\Delta t}{2}) = \dot{R}_i(t - \frac{\Delta t}{2}) + \ddot{R}(t)\Delta t \quad (1-76)$$

$$R_i(t + \Delta t) = R_i(t) + \dot{R}_i(t + \frac{\Delta t}{2})\Delta t \quad (1-77)$$

The advantage of the leap-frog algorithm is that both the position and the velocities are present and there is no subtraction type of error. The only disadvantage is that at given time step the velocities and the position are not available simultaneously.

1.4.1 Simulation protocol

Molecular dynamics follow the next five steps:

1. Establish the initial configuration of the system. It can be obtained by experimental data, a theoretical model or both. The positions of the atoms have to be chosen so that there are no atoms that overlap. To avoid that in future, a simple optimisation is performed.
2. Assign the initial velocities. It is done randomly by selecting the velocities from the Maxwell-Boltzmann distribution at a given temperature.
3. Equilibration of the system. The simulation continues until a steady state is reached and there are no big fluctuations in the selected variables.
4. Production run. After the system was equilibrated the needed measurements of the properties of interest are collected.
5. Data analysis. When the information is available, the final values of the properties can be calculated and compared.

1.4.2 Statistical mechanics

During the MD simulations different properties such as the pressure and the heat capacity of the system are determined through the position and momenta of N number of particles. $6N$ -dimensional space defined by the position and the momentum coordinates is called “phase space” of the system. At any time a particle occupies a point in this space:

$$X = (r_1, r_2, r_3, \dots, r_N, p_1, p_2, p_3, \dots, p_N) \quad (1-78)$$

where r_i is the position and p_i is the linear momentum coordinates.

Using different ensembles it is possible to derive all the states with different positions and momenta for a system keeping some of system's properties constant.

1.4.2.1 Microcanonical (NVE) ensemble

In the microcanonical ensemble the number of particles N , the volume V and the energies remain constant during the simulation. This is the most common ensemble having in mind that the Newton's dynamics is energy-conserving. The total energy of the system E_{tot} is a sum of the kinetic energy E_T and the potential energy E_V :

$$E_{tot}(\dot{R}, R) = E_T(\dot{R}) + E_V(R) \quad (1-79)$$

The potential energy is a function of the positions R and the kinetic energy depends on the velocities \dot{R} of the system's particles.

1.4.2.2 Canonical (NVT) ensemble and Boltzmann factor

In the canonical ensemble the properties that are kept unchanged are the number of Particles N , the volume V and the temperature T . The probability density function ρ for the NVT ensemble is proportional to the Boltzmann factor:²⁹

$$\rho_{NVT}(X) \propto \exp(-\beta E(X)) \quad (1-80)$$

where E is the total energy of the system and $\beta = 1 / k_B T$.

In a system the ratio between two different states X and X' is as follows:

$$\frac{\rho_{NVT}(X)}{\rho_{NVT}(X')} = e^{-\Delta E(r,q)/k_B T} \quad (1-81)$$

where

$$\Delta E = E(X) - E(X') \quad (1-82)$$

the normalisation factor in the proportion of all the states over the phase space is nothing else but the total partition function Q_{NVT} :

$$\rho_{NVT}(X) = \frac{1}{(h^{3N})N!} \frac{\exp(-\beta E(X))}{Q_{NVT}} \quad (1-83)$$

where h is the Planck constant and $N!$ reflects the indistinguishability of the N particles.

1.4.2.3 Isothermal-isobatic (NPT) ensemble

In the NPT ensemble the number of particles, the pressure P and the temperature T are kept constant. This allows a change in the volume of the simulated box.

1.4.2.4 Temperature coupling

The temperature coupling of the system is accomplished by a simulation of an exchange of heat with the so-called “external bath”.

1.4.2.4.1 Berendsen temperature coupling

The Berendsen thermostat³² is based on the deviation of the temperature T of the system from the temperature T_0 of an external bath:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \quad (1-84)$$

in this equation τ is a time constant, which is the intervals of contact of the system with the external bath. The greater τ , the weaker is the coupling. This type of thermostat keeps the temperature constant, scaling the velocities with a factor λ :

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau} \left(\frac{T_0}{T} - 1 \right)} \quad (1-85)$$

The Berendsen thermostat keeps the overall temperature constant but does not maintain the fluctuations of the temperature canonical. For this purpose the Nosé-Hoover thermostat is generally used.

1.4.2.4.2 Nosé-Hoover thermostat

It is a good idea to use a better canonical thermostat once the system is relaxed. First Nosé³³ proposed an extended thermostat for canonical simulations. Later this thermostat was modified by Hoover.³⁴ The Hamiltonian in the Nosé-Hoover coupling is extended by introducing a thermal reservoir and a friction force, proportional the product of each particle's velocity and friction term ξ . The friction term ξ is a dynamic quantity and thus has its own equation of motion.

Hoover's equation of motion is as follows:

$$m_i \ddot{R} = F - \xi m_i \ddot{R}_i \quad (1-86)$$

and the equation of motion for the heat bath is:

$$\frac{d\xi}{dt} = \frac{1}{Q}(T - T_0) \quad (1-87)$$

where T and T_0 are the instant and the reference temperatures and Q is a constant that determines the coupling strength. The Nosé-Hoover thermostat conserves the energy of the extended system.

1.4.2.4.3 Pressure coupling

The same idea as in the temperature coupling can be applied in the pressure coupling- the system is coupled to an external barostat. The generally used barostats are the Berendsen and the extended Parrinello-Rahman barostat.

1.4.2.4.4 Berendsen barostat

The Berendsen algorithm³² rescales the coordinates and box vectors every step to reach a reference pressure P_0 of the external barostat:

$$\frac{dP}{dt} = \frac{P_o - P}{\tau_p} \quad (1-88)$$

in the equation τ_p is the coupling constant.

1.4.2.4.5 Parrinello-Rahman thermostat

As the Berendsen thermostat, the weak coupling of the Berendsen barostat does not give exact canonical ensemble. The Parrinello-Rahman barostat³⁵ is analogous to the Nosé-Hoover in the pressure coupling.

In the Parrinello-Rahman barostat, the box vectors are expressed by the matrix b , which obeys the matrix equation of motion:³⁶

$$\frac{db^2}{dt^2} = VW^{-1}b'^{-1}(P - P_{ref}) \quad (1-89)$$

where V is the volume of the box, W is the matrix that determines the coupling strength and P and P_{ref} are the instant and reference pressures.

1.4.3 Applications of Molecular Dynamics

Molecular dynamics serve different purposes in chemistry and the most common are stated below:³⁷

1. Conformational sampling: the temperature gives the system additional kinetic energy and makes possible overcoming barrier and exploring the potential energy

surface. The trajectory contains the information of the conformational changes in the system.

2. Explicitly time-dependent properties: the MD simulations have memory of the system through the trajectory and that gives the opportunity to investigate many time dependent properties as diffusion coefficients, vibrational spectra, etc.
3. Sampling from statistical-mechanical ensembles: MD can be used to calculate ensemble averages.

Chapter 2 Enzymatic halogenation

2.1 Halogenation in nature and halogenated natural products

Halogenated molecules are widespread in nature, numbering over 4000³⁸ known compounds. In nature halogenated compounds can be found in fungi, lichen, bacteria, marine organisms, insects, higher animals, volcanoes and sediments, etc. Volcanoes generally emit inorganic halogenated products such as hydrochloric acid (HCl) and hydrofluoric acid (HF). Gases emitted by the Santiago volcano in Guatemala include organohalogens such as $\text{CF}_2=\text{CF}_2$, $\text{CF}_3\text{CF}=\text{CF}_2$, CHF_2Cl , $\text{CClF}=\text{CF}_2$, CH_3Cl , CFCl_3 , $\text{CCl}_2=\text{CHCl}$ and $\text{CCl}_2\text{FCClF}_2$.³⁹ Their formation is probably due to the high-temperature and -pressure reactions between HCl and HF and the organocompounds from the fossil soil.

Halogenated fulvic acids have been found in groundwater samples that date from around 5200 years ago.³⁹ Trichloroacetic acid has been shown to exist in some environmental compartments, including bog water, snow, rain and soil.

During a forest fire, the burning of biomass produces a large quantity of CH_3Cl . Dioxins and polychlorinated dibenzofurans are present in small quantities and it is for this reason they cannot be isolated (Figure 2-1)

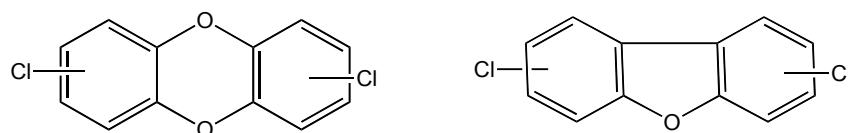


Figure 2-1. Chlorinated dioxins and benzofurans

Biosynthesised organohalogen compounds can be isolated from a variety of sources and may have diverse biological activity. Fungi and lichen living in symbiosis can yield different organocompounds. The lichen *Neophora laevigatum* synthesises **1** and falconensin E (**2**) from the soil fungus *Emericella falconensis* (Figure 2-2).

Falconensin E is an azaphilone, which inhibits diacylglycerol acyltransferase and acyl- CoA- cholesterol acyltransferase.³⁹

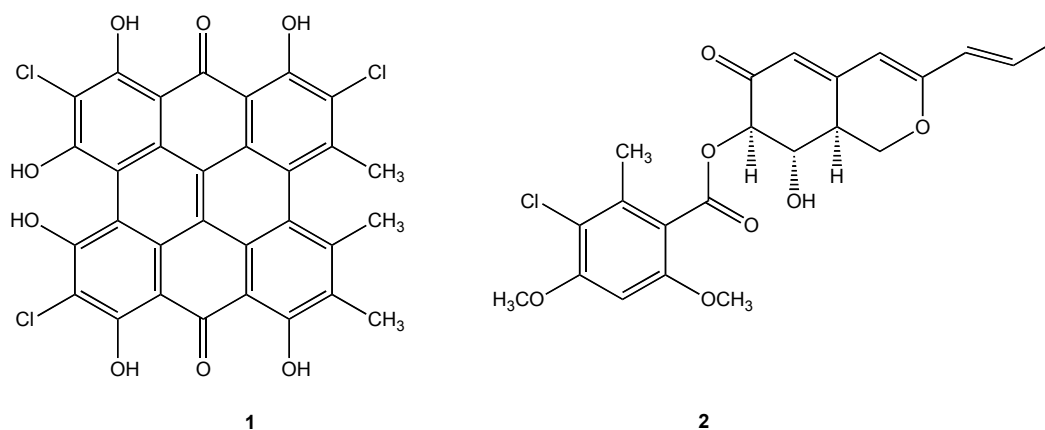


Figure 2-2. Chlorinated compounds isolated from lichen **(1)** and fungi (Falconensin E **(2)**)³⁹

Cryptophycins (**3**) are found in the alga *Nostoc sp.* GSV 224.⁴⁰ These compounds have strong antitumour activity. The first compound from this class that has been isolated is cryptophycin (Figure 2-3) and later 21 other analogues were isolated.

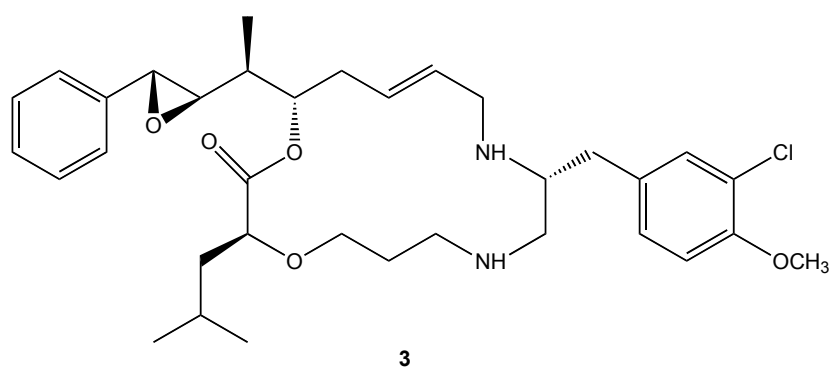


Figure 2-3. Chlorinated antitumor agent Cryptophycin-1 discovered in the alga *Nostoc sp.*³⁹

Streptomyces sp., bacteria synthesise the chlorinated hexapeptide Chloropeptin I (**4**), which inhibits the HIV- replication in the peripheral human lymphocytes. It could be put in combination with reverse transcriptase and protease inhibitors.⁴¹ Another bacterium- *Streptomyces venezuelae* produces the chlorinated antibiotic Chloramphenicol (**5**)(Figure 2-4).⁴²

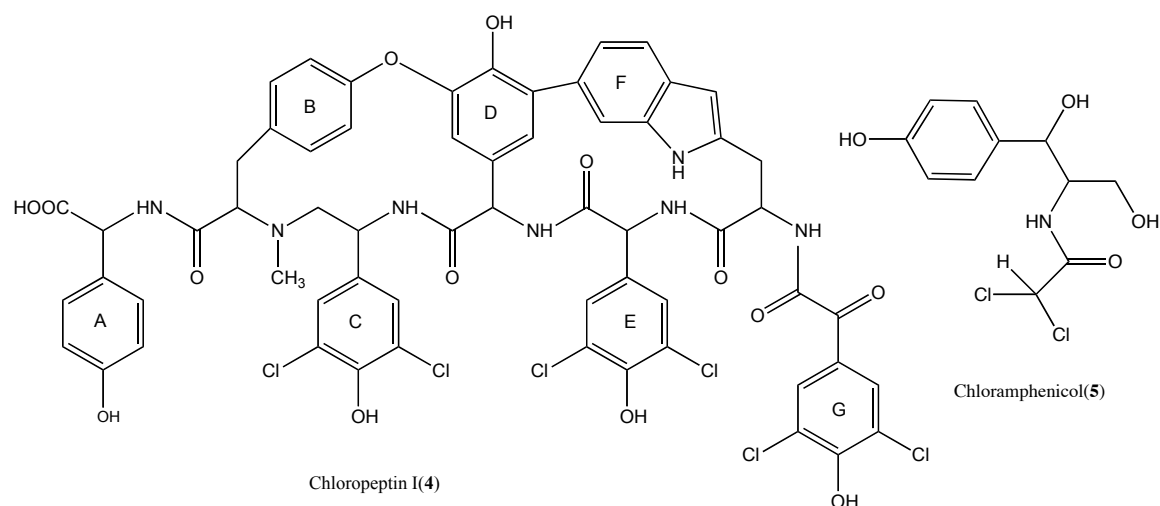


Figure 2-4. Organohalogens, produces by bacteria with HIV-inhibitor (**4**) and antibiotic (**5**) activities³⁹

Plants are a great source of halogenated compounds. Marine plants contain large quantities of organohalogens. *Plocamium telfairiae* synthesises telfairine (**6**), which is an insecticide, analogous to the industrial product lindane.³⁹ Monoterpene (**7**) from *Portiera hornemannii* has been shown to have anti-cancer activity (Figure 2-5).³⁹

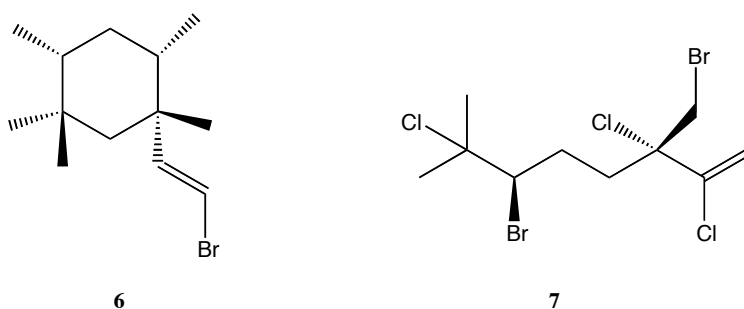


Figure 2-5. The organohalogens telfairine (**6**) and a monoterpene (**7**) isolated from inmarine organisms³⁹

In humans the only organohalogens produced are thyroid hormones (**8**) as well as recently isolated bromoester (**9**), which is purported to play a role in inducing REM (rapid-eye-movement) sleep (Figure 2-6).³⁹

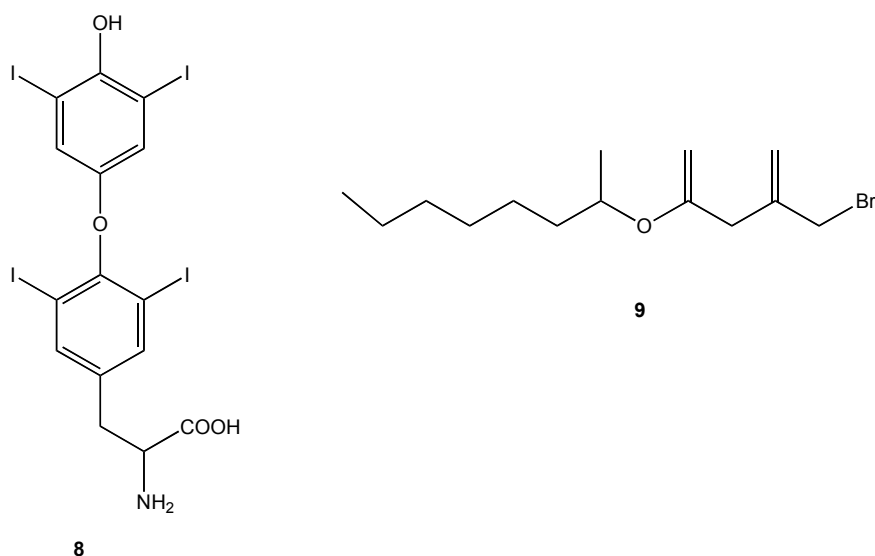


Figure 2-6. Organohalogens in human: the thyroid hormone thyroxine (**8**) and bromoester found in the cerebrospinal fluid and responsible for the REM sleep (**9**)³⁹

Chlorinated natural products typically predominate over brominated, iodinated and fluorinated ones. The type of halogen present in the natural product depends on the environment and the amount of halide present there. For example most bromides are synthesised from marine organisms, such as *Ptychodera flava laysanica*,³⁹ compared to chlorinated products, which can be found in terrestrial organisms. There are some mixed halogenated compounds that are known to exist in the literature. An example is a monoterpene in *Aplysia californica* with effective defence against predation activity.⁴³ The edible Hawaiian seaweed *Asparagopsis taxiformis* produces around 100 organohalogens.³⁹ Recently, first the fluorinase enzyme was identified in the bacterium *Streptomyces cattleya*,⁴⁴ which is involved in the production of a natural fluorinated compounds. The biological activity of the chlorinated molecules could be antitumor (rebeccamycin, cryptophycin), anti-HIV (chloropeptin I, ambigol A) or antibiotic (pyrrolnitrin, pyoluteorin, chloramphenicol).³⁸

2.2 Halogenation strategies

Classification of halogenating enzymes can be done, based either on the co-substrate of the reaction or the species of the halogenating donors. Halogenating enzymes can be divided according to the first classification into hydrogen peroxide (H_2O_2)-requiring haloperoxidases (haem-dependent or vanadium-dependent), the oxygen-dependent halogenases (flavin-dependent or non-haem iron-dependent) and the nucleophilic halogenases. The second category will be the one studied in this thesis and it includes enzymes that use hypohalite (X^+), the halogen radical (X^\bullet) or halide (X^-) as halogenating donors.

2.2.1 Halogenation by Hypohalite, XO^- (X^+)

The main type of halogenation is related to production of hypohalite (XO^-). The enzymes catalyze the two-electron oxidation of halide (X^-) to hypohalites (XO^-).⁴⁵ In this way the X^+ electrophile attacks electron-rich carbon atoms and does the halogenation. The two classes of enzymes, which could use hypohalites, are haloperoxidases and flavin-dependent halogenases.

2.2.1.1 Haem-Dependent haloperoxidases

Haem-dependent haloperoxidases are found in plants, fungi, bacteria and animals. Hager and co-workers discovered the first enzyme from that group in *Caldariomyces fumago*.⁴⁶⁻⁵¹ The enzyme catalyses the chlorination of late intermediate part of the biosynthesis of the natural antibiotic caldariomycin (Figure 2-7a). Thus the enzyme was named “chloroperoxidase”. The enzyme is able to catalyse reactions with the use of bromide and iodide but not fluoride. The characterization of chloroperoxidases was made using monochlorodimedone (Figure 2-7b), that is structurally similar to 2-chloro-1,3-cyclopentanedione (Figure 2-7c), an intermediate of the caldariomycin biosynthesis. The chlorination of monochlorodimedone can be monitored spectroscopically and used to identify the haloperoxidases in range of organisms.

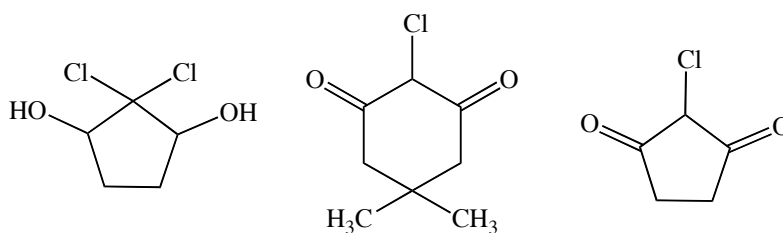
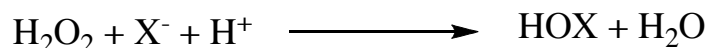


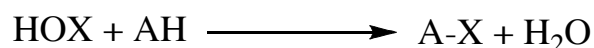
Figure 2-7. a) caldariomycin; b) monochlorodimedone; c) 2-chloro-1,3-cyclopentanedione⁴²

The general mechanism of heme-dependent haloperoxidases includes oxidation of chloride and bromide using hydrogen peroxide (H_2O_2). The hypohalous acid formed in the reaction leaves the active site of the enzyme and attacks the substrates activated for an electrophilic substitution in a nonenzymatic reaction:⁵⁰

Enzymatic reaction:



Nonenzymatic reaction:



$\text{X}=\text{Cl}, \text{Br}$ or I

Inside the enzyme the first reaction represents formation of HOX (Figure 2-8). First, the Fe^{III} is oxidised. A molecule of water is then eliminated and a Fe^{IV} -oxo compound is formed. A halide is then transformed into a hypohalite. The resultant hypohalite could be the halogenating agent, but it was proved that it provides a hypochlorous acid, which is the actual halogenating agent.⁵¹ Franssen and van der Plas⁴² could not measure the K_m (Michaelis-Menten constant) of the halogenation step of the pathway, which led them to believe that the chlorination step is done outside the enzyme, through formation of hypochlorous acid.⁴⁵ The haem-dependent haloperoxidases are not substrate specific or regioselective, which is another proof that the real halogenating agent is the hypohalous acid. Moreover, the crystallographic studies

revealed a possible binding site for hydrophobic substrates - when the concentration of substrate is smaller, the chlorination lowers. That could be explained just with chlorination in the enzyme.⁴⁹

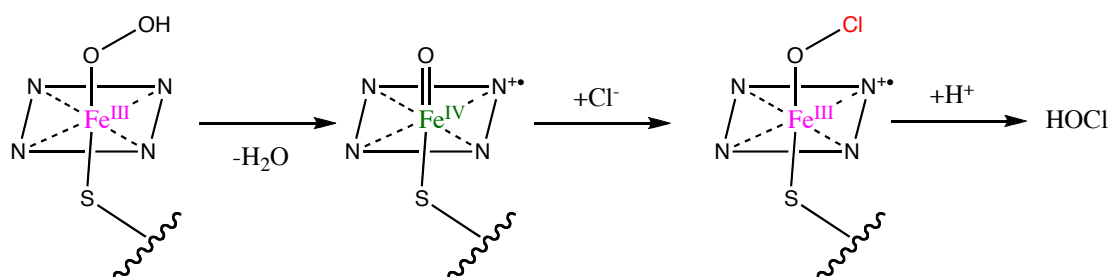


Figure 2-8. Formation hypochlorous acid by haem-dependent chloroperoxidases^{45,52}

2.2.1.2 Vanadium-dependent haloperoxidases

Most of the haloperoxidases have a haem co-factor, a specific subgroup of haloperoxidases exists, that are vanadium-dependent (Figure 2-9).^{46,47,49,50,53} Vanadium-dependent peroxidases were isolated from fungi and algae by Vilter.⁵⁴ The enzyme in the alga *Ascophyllum nodosum* was first identified as iodoperoxidase but later after reaction with monochlorodimedone was proven to actually be a bromoperoxidase.⁵⁵ Afterwards many vanadium-dependent haloperoxidases were found in marine algae, lichen and bacteria. In this class of enzymes the vanadium ion is bound to the imidazole ring of a histidine residue of the enzyme. The process of halogenation occurs in the same way as in the heme-dependent haloperoxidase, but without a change in the oxidation state of the vanadium ion. The product of halogenation is hypobromite.

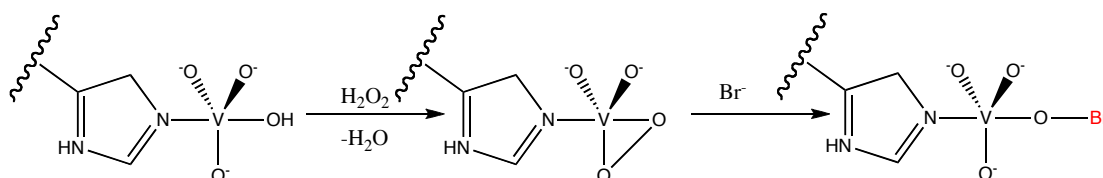


Figure 2-9. Formation of vanadium bound hypobromite by vanadium-dependent haloperoxidase^{45,52}

Experimentally these enzymes showed signs of regioselectivity or enantioselectivity. The first crystal structure of a vanadium-dependent haloperoxidase was solved for the chloroperoxidase from *Culvularia inaequalis*.⁵⁶ The vanadate-binding site is at the end of a ~10 Å solvent-accessible tunnel. It is proposed that this channel has an influence on the substrate specificity.

Vanadate bromoperoxidases seem to be important in the formation of marine natural products, e.g. in the halogenation and cyclisation of terpenes. After the discovery of such an enzyme in *Xanthoria parietina*, it was found that this group of enzymes exist in terrestrial species as well.

The nomenclature of the haloperoxidases is based on the most electronegative halide oxidised by the enzyme. Thus chloroperoxidases catalyse the oxidation of chloride, bromide and iodide, while bromoperoxidases catalyse only the oxidation of bromide and iodide. Hydrogen peroxide is not able to oxidise fluorine.

2.2.1.3 Flavin-dependent halogenases

Halogenation by hypohalite can also occur in the presence of FAD (flavin adenine dinucleotide), which is essential co-factor for the group of the flavin-dependent enzymes.^{45,48,49,53,57} There are two subclasses of flavin-dependent halogenases, one that catalyses the chlorination of small free molecules (RebH and PrnA) and the other that reacts with substrates attached to the thiolation domain in nonribosomal polypeptides synthase (NRPS) systems (PltA,⁵⁸ Sgc3⁵⁹). RebH catalyses the halogenation step in the biosynthesis of the antitumor agent rebeccamycin from *Lechevalieria aerocolonigenes*. PrnA and PrnC are both enzymes that participate in different stages of the pyrrolnitrin biosynthesis in *Pseudomonas fluorescens* B1915.⁶⁰ PyrH and Thal catalyse chlorination at positions 5 and 6 of tryptophan in the pyrroindomycin biosynthesis in *Streptomyces regosporus* and thienodolin biosynthesis in *Streptomyces albogriseolus*. PltA is a pyoluteorin biosynthesising enzyme in *Pseudomonas fluorescens* Pf-5 along with HalB that catalyses the pentachloropseudilin biosynthesis in *Actinoplanes* sp. ATCC 33002.⁶¹

Flavin-dependent enzymes are regioselective. The main substrate of their activity is tryptophan. Halogenations at positions 5, 6 and 7 in the indole ring are currently known (Figure 2-10). The general mechanism of halogenation includes FAD, O₂ and Cl⁻.

Enzyme	Organism	Substrate	Product	Structure	Comments
RebH	<i>Lechevalieria aerocolonigenes</i>	tryptophan	7-chloro-tryptophan	⁶²	-
PyrH	<i>Streptomyces regosporus</i>	tryptophan	5-chloro-tryptophan	⁶³	-
Thal	<i>Streptomyces albogriseolus</i>	tryptophan	6-chloro-tryptophan	-	-
PrnA	<i>Pseudomonas fluorescens</i>	tryptophan	7-chloro-tryptophan	⁶⁴	-
PrnC	<i>Pseudomonas fluorescens</i>	aminopyrrol nitrin	monodechloro aminopyrrol nitrin	-	-
PltA	<i>Pseudomonas fluorescens</i> Pf-5	Pyrolyl-S-PltL	5-chloropyrrolyl-S-PltL intermediate; 4,5-dichloro pyrrolyl-S-PltL product	-	NRPS, Double chlorination, requires PltL
HalB	<i>Actinoplanes sp.</i> ATCC 33002	monodechloro amino pyrrol nitrin	aminopyrrol nitrin	-	-
Sgc3	<i>Streptomyces globisporus</i>	β-tyrosyl-S-SgcC2	(S)-3-chloro-β-tyrosyl-S-SgcC2	-	NRPS; requires O ₂ , Sgc2 carrier

Table 2-1. FAD-dependent halogenases

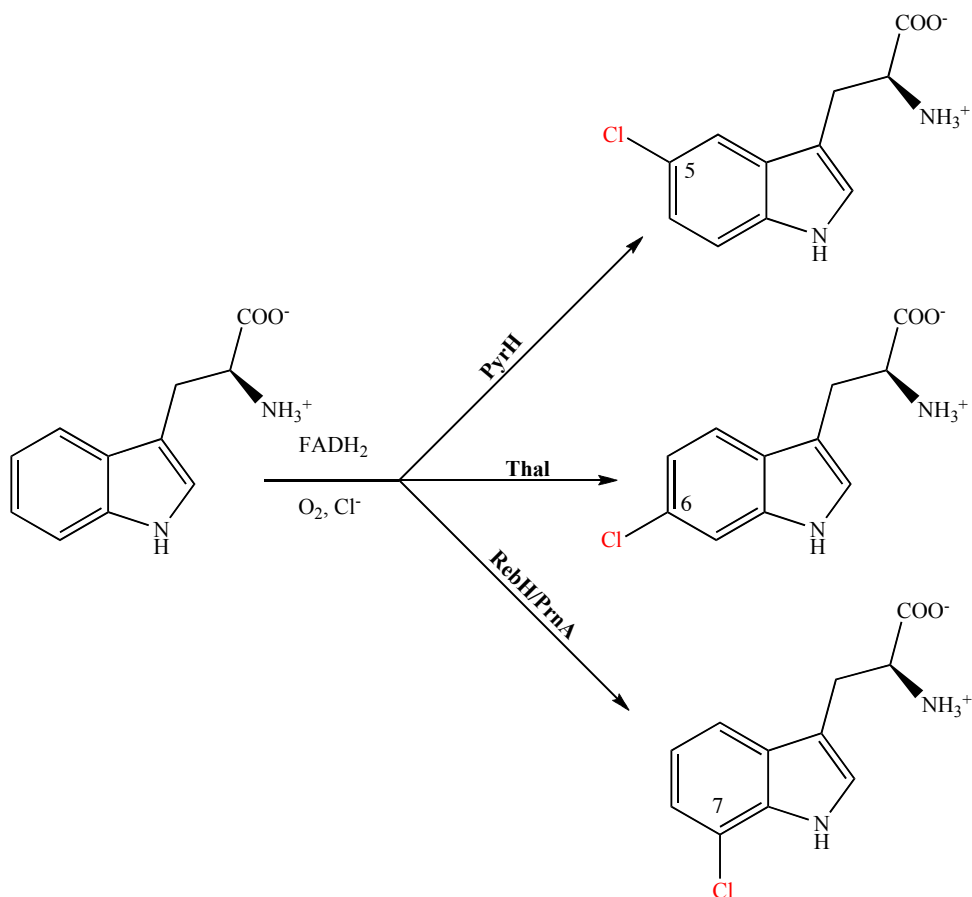


Figure 2-10. Enzymes from the flavin-dependent halogenase family, catalysing regiospecific chlorinations^{49,63}

2.2.1.3.1 PyrH

The crystal structures of two native enzymes and one mutant E46Q of PyrH have been solved.^{63,65} The resolutions of the three structures are 2.4 Å, 1.7 Å and 2.5 Å respectively. In the crystal structure PyrH is a dimer of dimers. PyrH is a dimer like PrnA and RebH with 40% sequence identity with PrnA, which is due mainly to similarities in the FAD binding domain (Figure 2-11). Tryptophan is situated at 9 Å distance from the FAD. The residues of primary importance for the mechanism of chlorination in PrnA- K79 and E346 are conserved in PyrH as K75 and E354 (Figure 2-12). Overall the secondary structure of PyrH remains without major changes compared to PrnA and RebH but the difference comes from the connecting loops. Among all the tryptophan flavin-dependent halogenases PyrH is the only one with C5 regioselectivity because of shielding of the other possible substitution positions in

tryptophan. PyrH binds its substrate in such a way that C5 is in the best position to react with the key residues (glutamate and lysine) proposed to participate in the mechanism of the reaction.

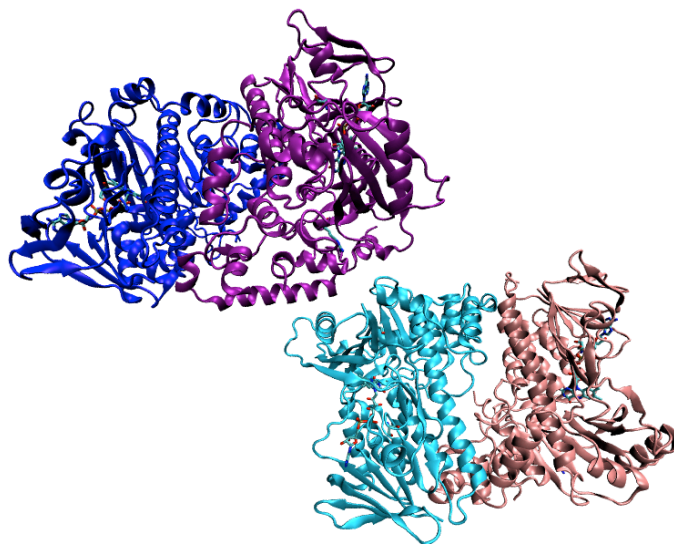


Figure 2-11. PyrH dimer structure downloaded from PDB (2WET.pdb)⁶⁶

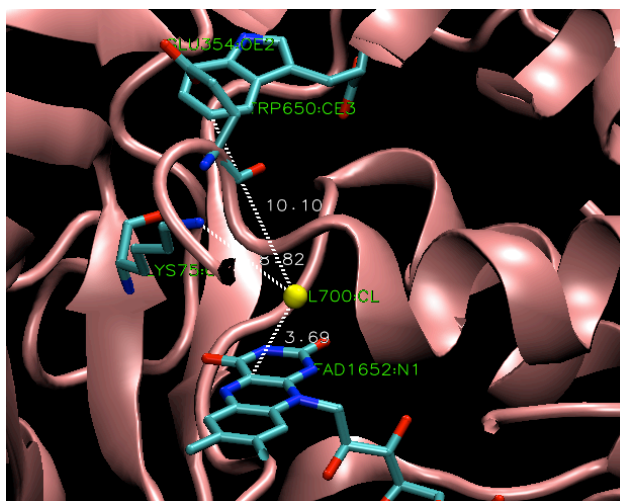


Figure 2-12. Crystal structure of the active centre of PyrH with Cl^- , FAD and tryptophan bound, downloaded from PDB (2WET.pdb)⁶⁶

2.2.1.3.2 RebH

RebH catalyses the chlorination step in the pathway of the biosynthesis of the indocarbazole natural product rebeccamycin (Figure 2-14).⁶⁷ It is a dimer with structure similar to PyrH and PrnA. It has been isolated from *Lechevalieria*

aerocolonigenes. The structures of apo-RebH and RebH have been solved with resolutions of 2.5 Å and 2.15 Å (Figure 2-13).⁶² The RebH structure has bound FAD and chloride. RebH shows 55 % sequence identity and 0.68 Å root mean square deviation (rmsd) compared to PrnA. The dimer in solution is stabilised by 14 hydrogen bonds and a salt bridge between Arg387 and Glu432 residues. The differences in the structures of RebH and PrnA are: 1) A solvent-exposed loop in RebH (86-105) is expanded by eight residues compared to PrnA (87-97); 2) A difference in conformation of a loop that in PrnA forms hydrogen bonds with the FAD co-factor and with waters connected to FAD, but adopts an *open* conformation in RebH without any connections to FAD, making the co-factor binding site accessible to water. Difference exists between apo-RebH and RebH and it is consistent with changes in the tryptophan-binding site (residues 111-114). The tryptophan binding changes the position of the carbonyl group of Gly112 and moves the side chain of Leu113 5 Å.⁶² The loop becomes more rigid, which reflects in well-defined charge density of that loop. The result of these structure rearrangements is that the substrate and water stabilising its position, become trapped inside the enzyme. The enzyme binds reduced FAD (FADH₂), oxygen, chloride (Cl⁻) and tryptophan. Halogenation by this enzyme requires the presence of RebF (Flavin reductase), which catalyses the NADH-dependent reduction of FAD to FADH₂.^{64,68,69}

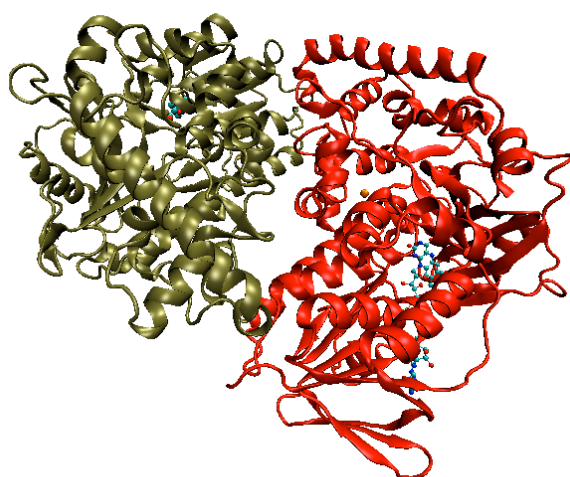


Figure 2-13. RebH structure with bound FAD and Cl⁻, downloaded from PDB (2OAL.pdb)⁶⁶

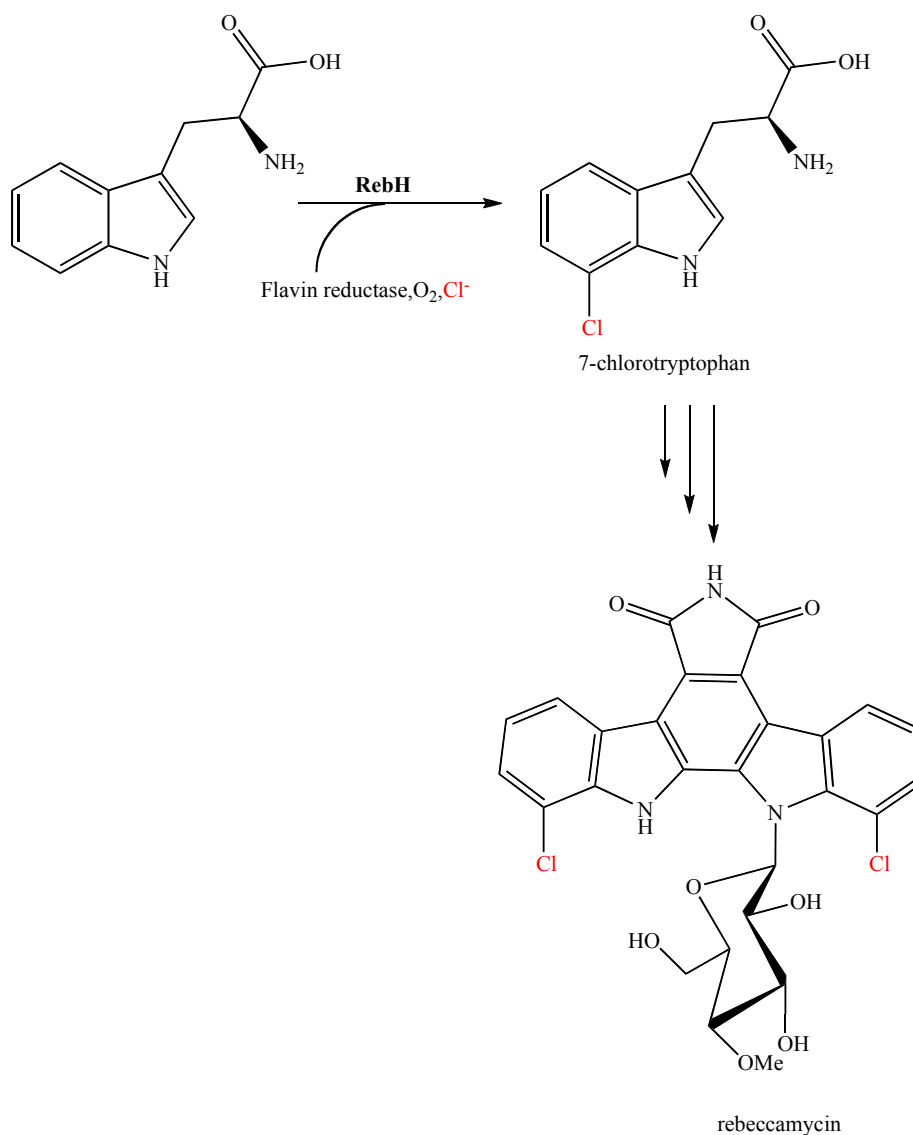


Figure 2-14. Formation of 7-chlorotryptophan as a first step of the biosynthesis of rebeccamycin by RebH⁶⁸

2.2.1.3.3 PrnA

PrnA is an enzyme with a dimer structure in solution and has 538 amino acid residues per chain of the dimer (Figure 2-15). The enzyme takes part in the biosynthesis of the natural antifungal agent pyrrolnitrin, produced by the bacterium *Pseudomonas fluorescens*. The PrnA structure was solved with 1.95 Å resolution.⁶⁴ The structure includes residues from 2 to 518. It contains two domains; one is the FAD-containing domain and another that binds the tryptophan molecule. The flavin-binding module is constructed by β -sheets and α -helices. The substrate-binding site of the enzyme

contains just α -helices.⁴⁶ The chloride is bound in a pocket facing isoalloxazine ring of FAD. In this pocket the chloride is solvent-protected from the FAD co-factor and it does not have real contact with it. The tryptophan-binding domain is 10 Å from the FAD co-factor. A channel connects the cofactor and the substrate-binding sites.^{52,70-75}

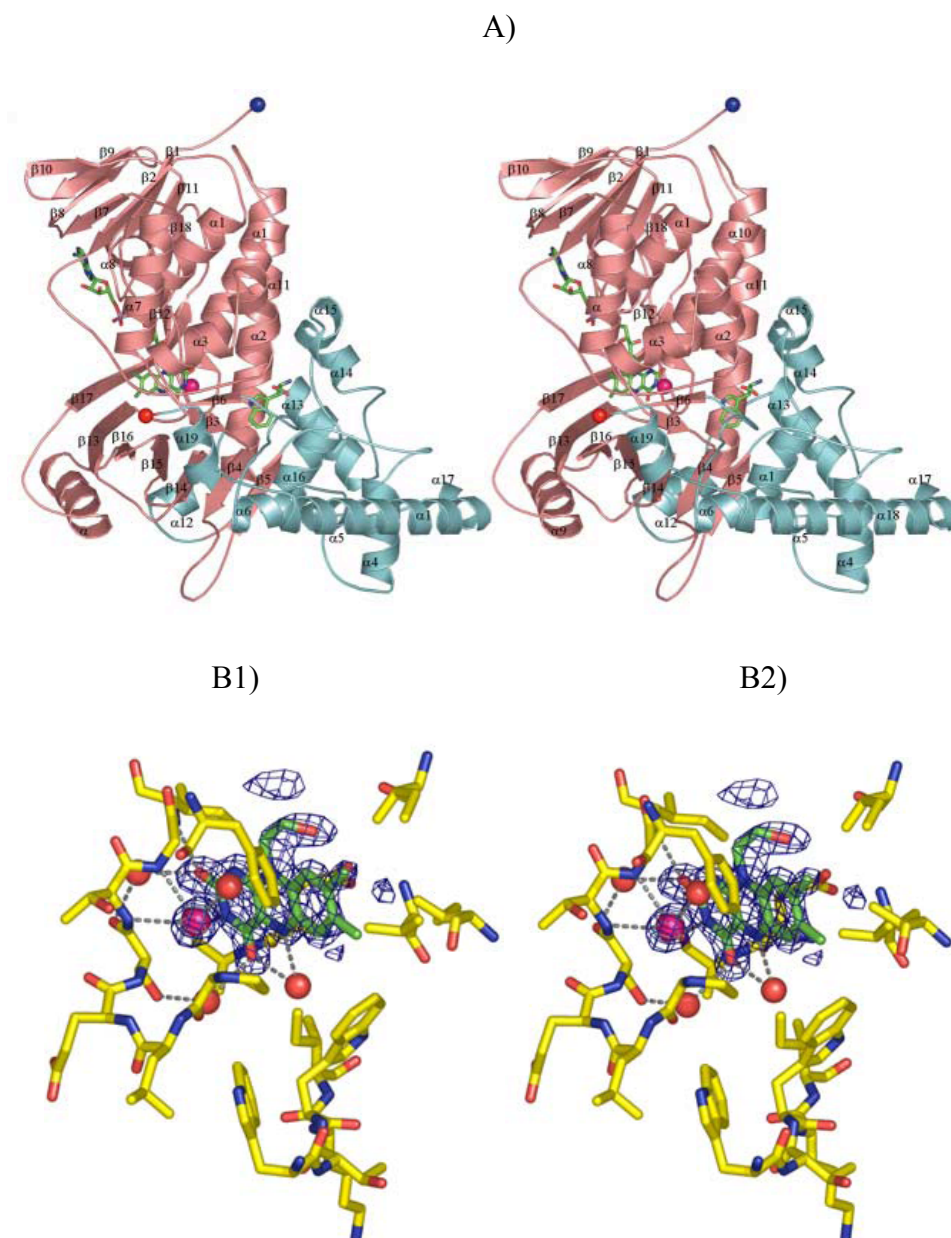


Figure 2-15. Stereo pictures of the PrnA monomer. A) The N-terminus of the enzyme is coloured in blue and the C-terminus in red, the chloride is a pink sphere, FAD and 7-chlorotryptophan are represented in sticks. B) The FAD and chloride (B1) and the substrate-binding (B2) sites of the enzyme are shown in sticks. The hydrogen bonds are shown as dotted lines⁶⁴

PrnA catalyzes the first step of the biosynthesis of pyrrolnitrin- the chlorination of tryptophan to 7-chlorotryptophan. The other steps of the pathway are catalysed by PrnB, PrnC and PrnD (Figure 2-16). PrnC is a chlorinating enzyme as well, but it belongs to a different group of halogenases. It catalyses the chlorination of pyrrol ring in monochloroaminopyrrolnitrin in another step further along the pathway. The analysis of the gene clusters encoding the PrnA and PrnC enzymes showed strong homologies in both enzyme gene sequences.³⁸ They have weak similarities with the flavin-containing monooxygenases such as para-hydroxybenzoate hydroxylase (PHBH).⁶⁴ The monooxygenase family of enzymes uses reduced FADH₂ to bind oxygen to form a hydroperoxyflavin reactant (FAD-4 α -HOOH) like the 7-chlorotryptophan halogenases. The difference between the monooxygenases and PrnA is that the monooxygenases do not have a separated substrate-binding domain in their structure. The investigation of the chlorinating activity PrnA and PrnC *in vitro* in a *P. fluorescens* mutant clarified the difference to chloroperoxidases in requiring NADH, O₂ and chloride (but not H₂O₂) to regioselectively produce 7-chlorotryptophan and aminopyrrolnitrin, respectively.⁵⁷

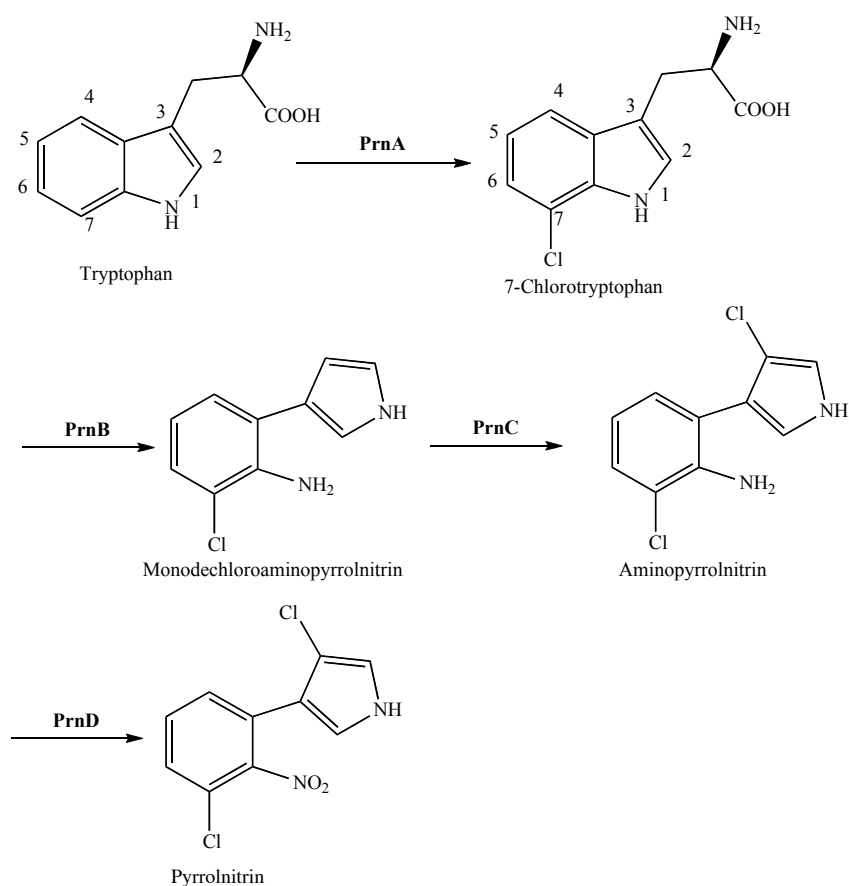


Figure 2-16. Biosynthesis of the natural antibiotic pyrrolnitrin in *P. fluorescens*⁴²

The formation of 7-chlorotryptophan occurs in several steps (Figure 2-17). The enzyme needs FADH_2 as a co-factor for the reaction. For this purpose, FAD is reduced to FADH_2 by flavin reductase with the participation of NADH (reduced nicotinate dinucleotide). The reduced cofactor binds oxygen to form an oxidised compound ($\text{FAD-4}\alpha\text{-HOOH}$). A hypothesis that was later rejected suggested a direct reaction with the substrate to form an epoxide. The oxygen atom in the arene oxide would be substituted by chloride (Cl^-), leading to formation of the product. This mechanism was rejected after experimental kinetics studies showed no direct interaction between tryptophan and $\text{FAD-4}\alpha\text{-HOOH}$. The latest mechanistic suggestions include the formation of hypochlorous acid (HOCl) after an attack of a chloride (Cl^-) on the hydroperoxyflavin. The acid could then directly react with tryptophan, form a hydrogen bond with a lysine residue in the enzyme; or produce a chloramine after a reaction with this residue. Chloramine is a strong electrophile and an electrophilic aromatic substitution could occur between it and the tryptophan molecule. The primary product of the reaction is a σ -complex of 7-chlorotryptophan. After a proton elimination 7-chlorotryptophan is formed.

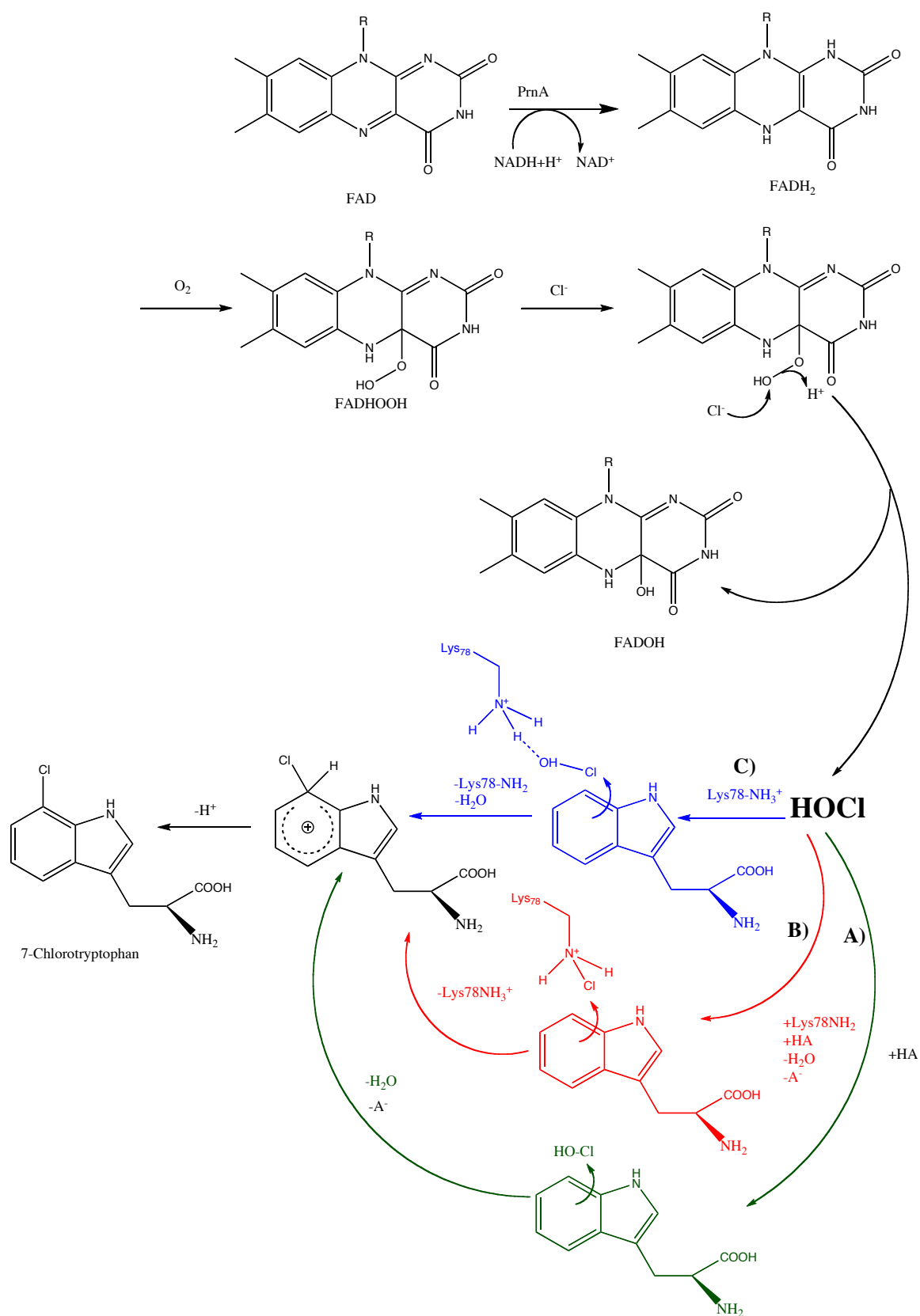


Figure 2-17. Formation of 7-chlorotryptophan: A) Chlorination by the free HOCl; B) Walsh and co-workers mechanism of chlorination through a formation of chloramine with the participation of Lysine78;⁶⁸ C) Naismith et *al.* mechanism of chlorination with hydrogen bond HOCl to the Lysine78 residue⁶⁴

2.2.2 Halogenation via Halogen, X[•]

The activity of a non-haem iron-dependent halogenase was first encountered by Vaillancourt in 2005.⁷⁶ This group of enzymes uses O₂ and α -ketoglutarate to incorporate chloride at inactivated carbon centres. They contain non-haem iron. Representative of this class is the SyrB1/SyrB2 halogenases in *Pseudomonas syringae*, responsible for the production of the liponapeptidolactone syringomycin E. SyrB2 catalyses the chlorination of the γ -methyl group of L-threonine but does not accept free L-threonine as a substrate. Threonine has to be bound to the phosphopantethenyl group of the thiolation domain of SyrB1 (Figure 2-18). The reaction mechanism suggests the formation of a substrate radical by abstraction of an H-radical by the Fe^{IV}=O intermediate. The substrate radical recombines with a Cl from the Fe^{III}-OH intermediate (Figure 2-19). A similar enzyme is CytC3, which adds two chlorine atoms to L-aminobutyric acid, which is attached to the thiolation domain of CytC2 in the biosynthesis of amentomycin by *Streptomyces sp.* RK95-74. The non-haem iron halogenases also catalyse the formation of cyclopropane-containing bacterial metabolites. An example of that is the formation of coronamic acid by CmaC in *Pseudomonas syringae*.⁵⁰

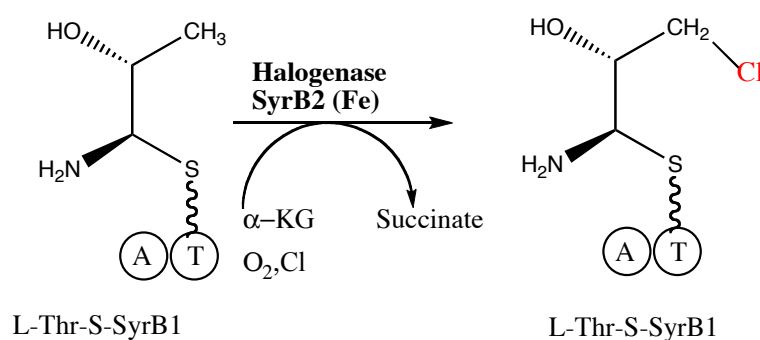


Figure 2-18. Non-haem dependent halogenase in which previous to the halogenation the substrate has to be bound to the thiolation domain of the SyrB1 carrier⁵⁰

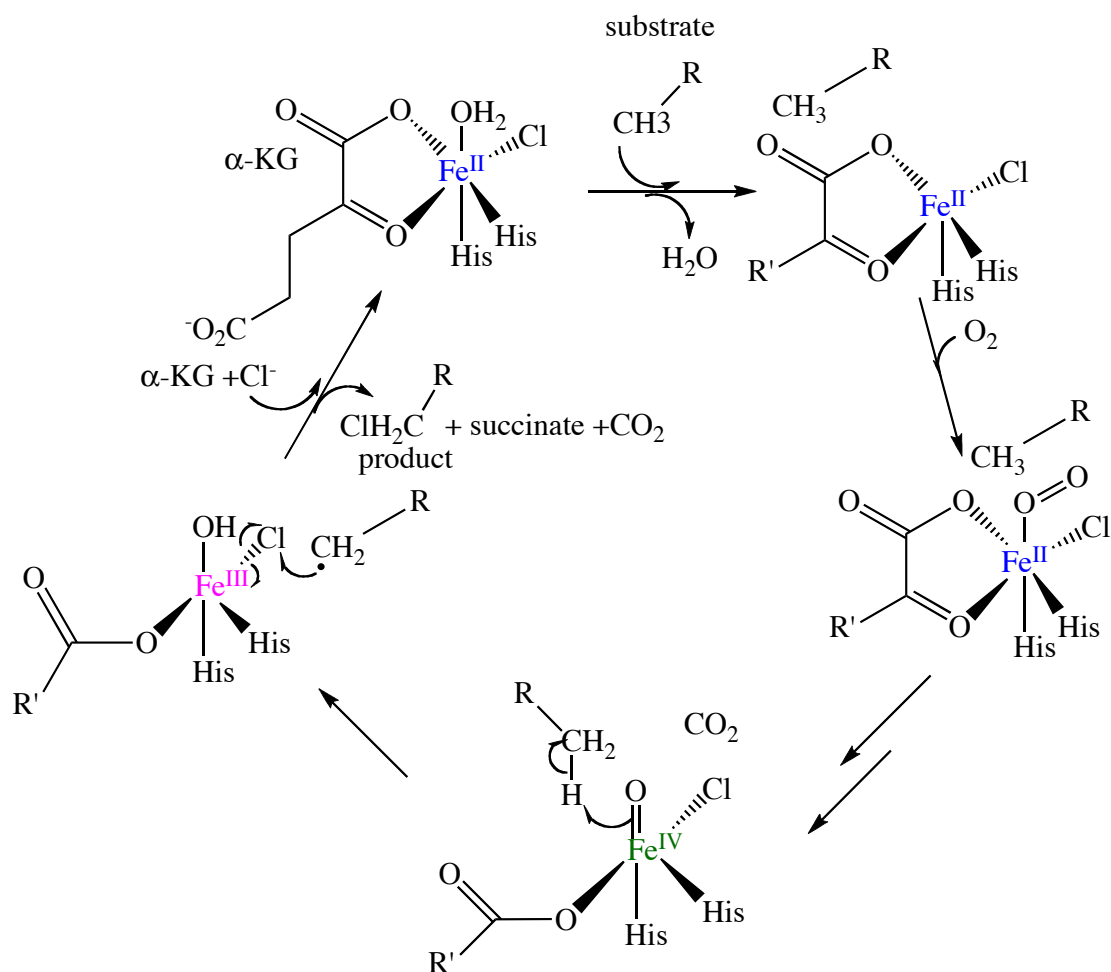


Figure 2-19. Mechanism of halogenation by X^- of non-haeme iron O_2^- and α -ketoglutarate-dependent halogenases⁵²

SAM is involved in the formation of monohalomethane but in this case as a methyl donor. The first methyltransferase was discovered by Wuosomaa and Hager⁷⁷ and isolated from the rot fungus *Phellinus promaceus*.

Non-enzymatic reactions of transfer of halides ions to organic compounds were detected as a part of the biosynthesis of cyanosporasides A and B from *Salinispora tropica*.⁴⁶

2.2.3 Halogenation via Halide, X^-

A pathway with a halide ion acting as a nucleophile is the fluoroacetate biosynthesis by *Streptomyces cattleya*⁵⁷. The enzyme catalysing the reaction is the 5'-fluoro-5'-

deoxyadenosine synthase (5'-FDAS) or the so-called fluorinase, whose structure was solved in 2004. Two conditions have to be satisfied for the reaction to be possible. Firstly, F^- must be stabilised and solvated inside the active site; secondly, an electrophilic carbon must be present in the organic co-substrate that F^- can attack. The stabilisation is provided by a serine side chain in the active site through a possible hydrogen bond and the electrophilic co-substrate is S-adenosylmethionine (SAM). The mechanism is consistent with S_N2 substitution of SAM. 5'-fluoro-5'-deoxyadenosine (5'-FDA) is formed as an intermediate and metabolised afterwards to 4-fluorothreonine and fluoroacetate via fluoroacetaldehyde (Figure 2-20).⁵⁰

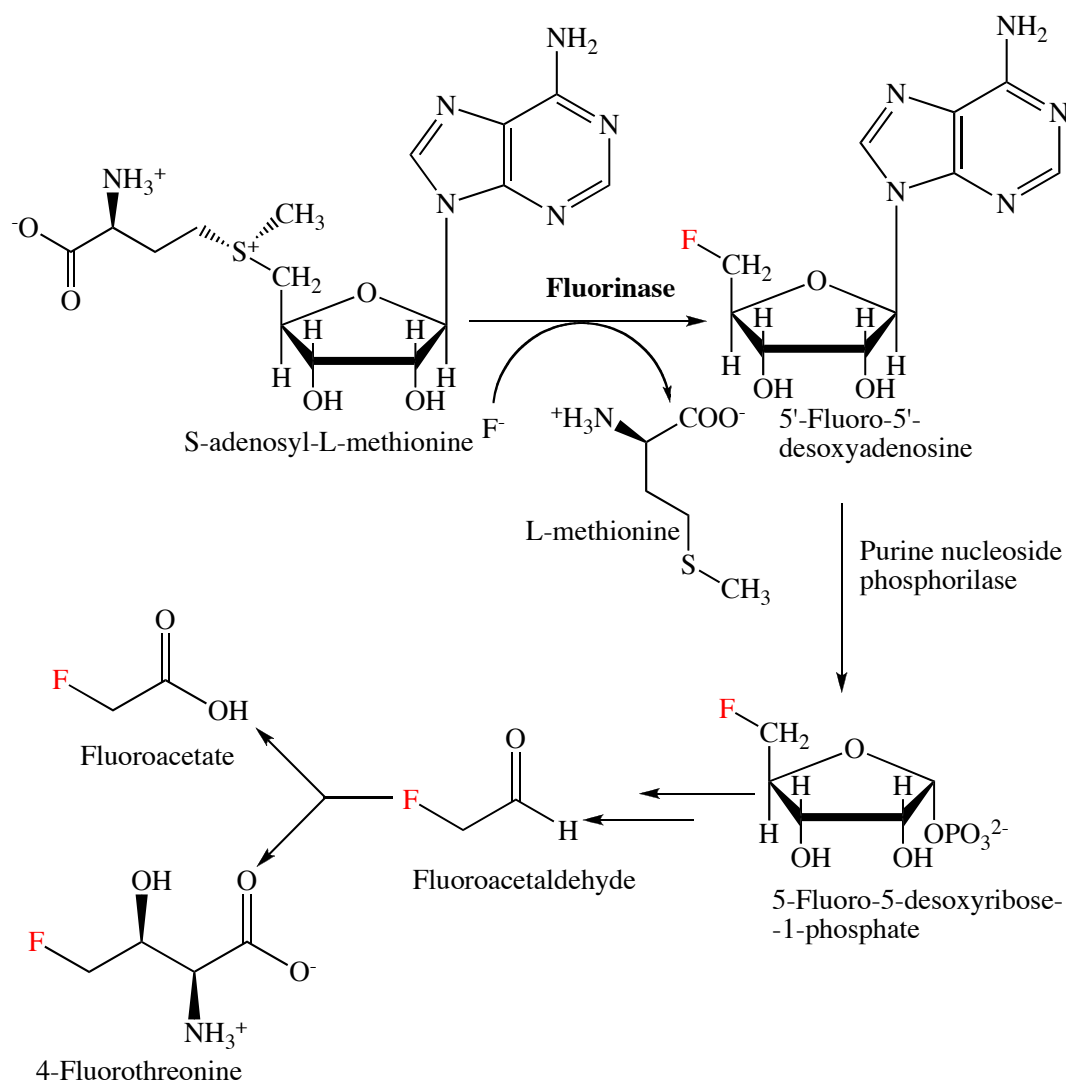


Figure 2-20. Fluorinase mechanism of halogenation involving transformation of F^- and SAM to 5'-FDA and L-methionine and further metabolism of 5'-FDA to fluoroacetate and 4-fluorothreonine⁵⁰

2.3 Electrophilic aromatic substitution

An electrophile can be a positively charged ion. Three groups of electrophiles in electrophilic aromatic substitution exist: 1) Electrophiles taking part in both activated- and deactivated-aromatic ring substitutions; 2) electrophiles taking part in activated-aromatic ring substitution; and 3) electrophiles participating in aromatic substitution with strongly activated aromatic rings. The mechanism of electrophilic aromatic substitution involves several steps. A formation of π - and σ -complexes is included at some stage in the reaction pathway. The formation of a π -complex does not necessarily occur. It is a donor-acceptor type of complex and its formation is quite reversible. The electron donor is the aromatic ring. The electrophile in this case is situated perpendicular to the aromatic ring. The next stage of the reaction is the formation of a σ -complex, which is a cationic intermediate. The carbon is bound to the electrophile and the hydrogen at the same time. The positive charge is delocalised over the whole molecule (Figure 2-21). The formation of the σ -complex may be reversible as well. The way of the reaction depends on the chosen electrophile. The formation of the σ -complex is usually the rate-determining step of the reaction. The last step in the reaction is a proton extraction. Before this final step a π -complex can be again formed as an intermediate.⁷⁸

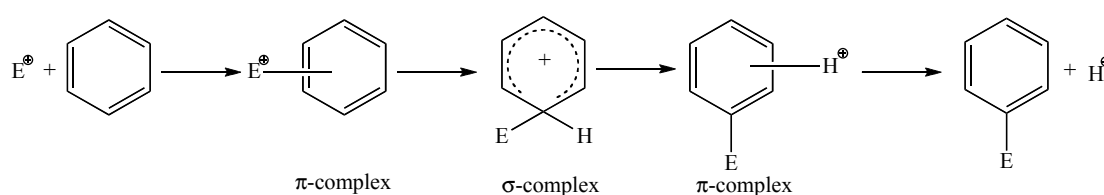


Figure 2-21. Mechanism of electrophilic aromatic substitution

2.4 Possible electrophiles involved in the tryptophan chlorination at position C7

In our electrophilic aromatic substitution the possible chlorinating agents for producing 7-chlorotryptophan are free hypochlorous (HOCl) acid; HOCl hydrogen bonded to the Lys78 residue of the enzyme or the chloramine (-NHCl) formed from the Lys78 and the hypochlorous acid.

Hydroperoxyflavin (FADHOOH) reacts with chloride (Cl^-) to form the hypochlorous acid (Figure 2-22). The acid could remain free or react with the lysine residue in the enzyme. It could possibly form a hydrogen bond with Lys78 and in this case the acid is positioned so as preferentially to react with the tryptophan at the 7-position (Figure 2-23). Alternatively a chloramine could be produced from the reaction between the acid and the enzyme residue. The chloramine can exist either protonated (Figure 2-24) or non-protonated (Figure 2-25).

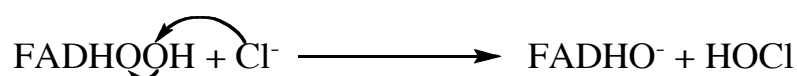


Figure 2-22. Formation of hypochlorous acid

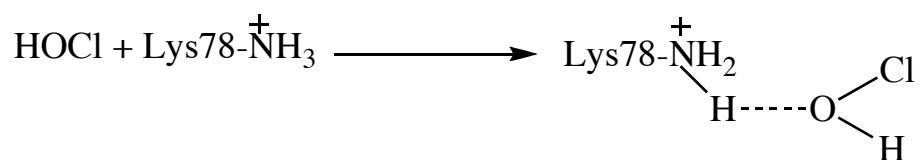


Figure 2-23. Hydrogen bond between the hypochlorous acid and the Lys78-residue

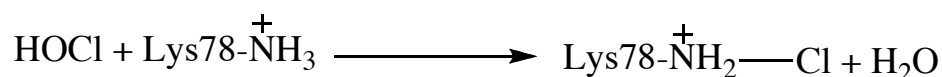


Figure 2-24. Reaction between protonated Lys78-residue and the hypochlorous acid

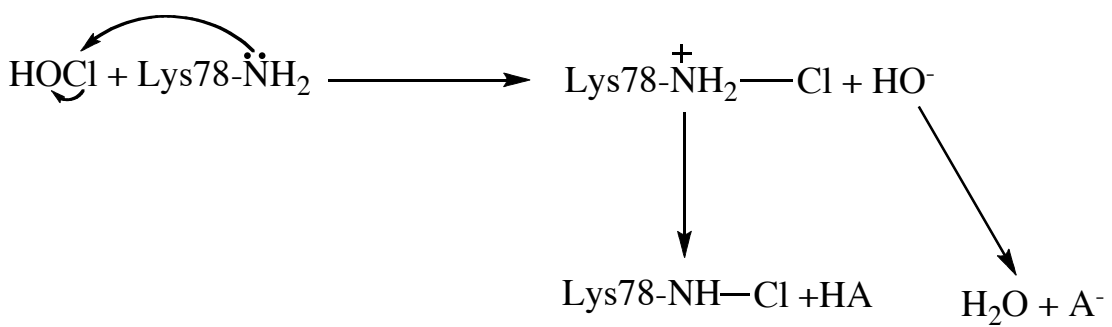


Figure 2-25. Reaction between the hypochlorous acid non-protonated Lys78

In Chapter 3 will be discussed the results of the calculations considering the possible mechanisms of the reaction.

Chapter 3 QM calculations on PrnA

3.1 Introduction

The previous chapter gave an overview of the different halogenating enzymes and in particular, an insight into the PrnA halogenating system. In this chapter a closer look will be carried of the active centre of PrnA, the reactions occurring there, the transition states and the products of the enzymatic activity of PrnA. In the active centre, the co-factor FAD, the chloride anion, the substrate tryptophan (ZTRP) and the Lys78 residue play the main role (Figure 3-1).

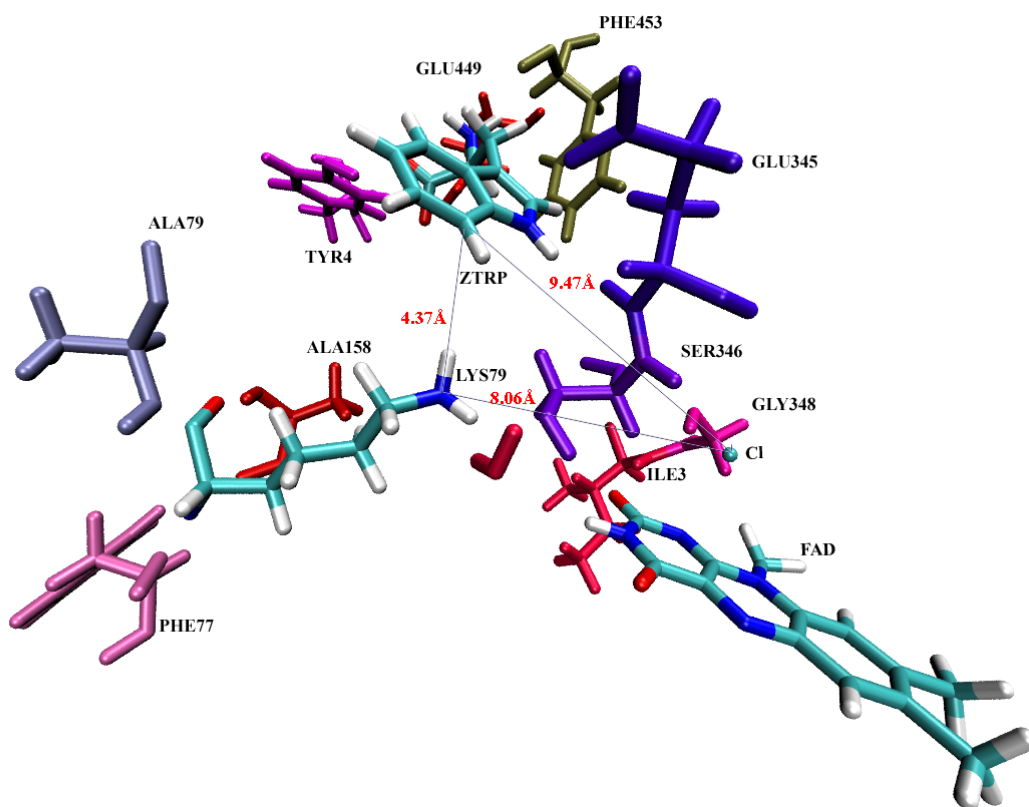


Figure 3-1. Active centre in the crystal structure of PrnA

As we discussed in Chapter II, the tryptophan-chlorinating agent can be the free hypochlorous acid, the acid, hydrogen bonded to residue Lys78, or the chloramine, formed from the Lys78 and acid (Figure 3-2). The Lys78 residue is situated in the channel, connecting FAD and tryptophan. The Lys78 residue is a weak base and can adopt either the protonated or non-protonated form.

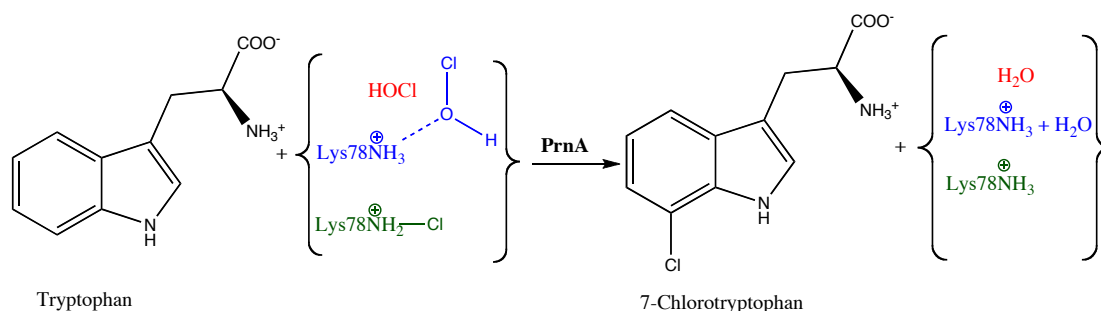


Figure 3-2. Formation of 7-Chlorotryptophan from the attack of different halogenating agents on tryptophan

Models for the QM-calculations were defined. Indole, (chloroammonio)methane and hydrogen peroxide were used as computational models instead of tryptophan, Lys78-NH₂Cl⁺ and FADHOOH (Figure 3-3).

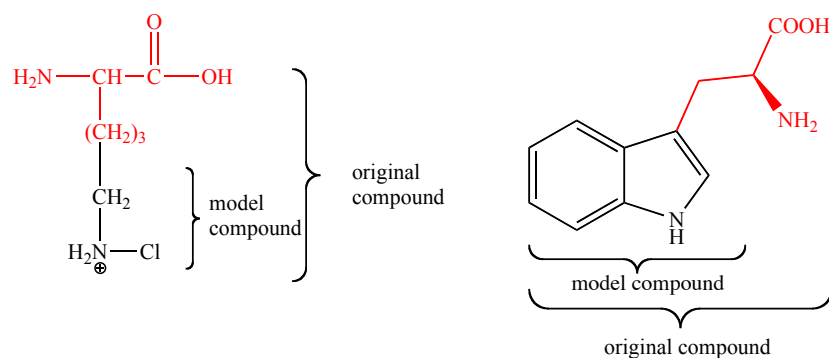


Figure 3-3. The model compounds used for the reaction of chlorination of tryptophan to 7-chlorotryptophan occurring in the active centre of the PrnA

The reactions in the active centre were investigated at QM level. Reaction energies, along with activation energies with various halogenating agents, were calculated. The

regioselectivities and stabilities of the products of the chlorination, catalysed by PrnA will be investigated as well.

3.2 Computational methodology

3.2.1 Computational models and augmented basis set

The *TURBOMOLE* program package⁷⁹⁻⁸³ was used for DFT calculations with the standard *TURBOMOLE*'s optimizer *Relax*,⁸⁴ running a *Jobex* script. The optimisations were performed in internal redundant coordinates. The calculations were carried out with the TPSS (meta-GGA functional of Tao, Perdew, Staroverov and Scuseria) functional,⁸⁵ employing the def2-TZVP (triple ζ -valence polarization) basis set and def2-TZVP auxiliary basis set.⁸⁶ The basis set was augmented by diffuse functions of s- and p-type on all the atoms to def2-TZVP+ level (Table 3-1).

Atom	Exponent	
	s	p
H	3.2466181×10^{-2}	-
C	3.9421656×10^{-2}	3.5011523×10^{-2}
N	5.5356409×10^{-2}	4.8837750×10^{-2}
O	7.3678273×10^{-2}	5.7711255×10^{-2}
Cl	6.4237604×10^{-2}	4.4417963×10^{-2}

Table 3-1. Exponents added to the diffused functions in the def2-TZVP+ basis set

3.2.2 COSMO (Conductor-like Screening Model) calculations

The QM-COSMO model,⁸⁷ and experimental data, were used for validating the solvation radii of the elements, from which the investigated compounds were built. COSMO is a continuum solvent model that builds a cavity around the compounds.

The surrounding space has a dielectric constant equal to that of water, $\epsilon_r = 78.4$. The compounds were optimized in a gas phase, and in a solvent, in order to obtain the solvation energies, which represent the difference between both energies.

$$G_{solvation} = G_{solvent} - G_{gas} \quad (3-1)$$

The computed solvation energies were compared to the experimental data, for calibration. The investigated default radii for the atoms in the structures were as follows: Cl= 2.05 Å, O= 1.72 Å, N= 1.83 Å (Table 3-2).

Compound	Calculated solv. Energy in kJ/mol	Experimental solv. Energy in kJ/mol ⁸⁸	Deviation(ΔG)
Chloride anion	-325.28	-311.70	-13.58
Hydroxide anion	-407.60	-438.06	30.47
Methylamine	-14.55	-19.10	4.55
Methylammonium	-302.57	-319.70	17.13
Chlorobenzene	-12.98	-4.69	-8.29
Pyridine	-21.96	-19.70	-2.26
Water	-33.54	-26.40	-7.14

Table 3-2 Calculated and experimental solvation free energies and deviations

The results showed that it was needed a reparameterization of the atom radii, comparing the calculated solvation energies to the experimental ones. We tried to obtain energies closer to the experimental values by altering the radii.

Compounds with Chlorine	Radius (Cl)/ Å	Calculated solv.energy in kJ/mol	Experimental solv.energy in kJ/mol	Deviation (ΔG)
Chloride anion	2.10	-318.15	-311.7	-6.45
Chlorobenzene	2.10	-12.78	-4.69	-8.09
Chloride anion	2.15	-311.34	-311.70	0.36
Chlorobenzene	2.15	-12.60	-4.69	-7.91
Chloride anion	2.20	-304.79	-311.70	6.91
Chlorobenzene	2.20	-12.51	-4.69	-7.82

Compounds with Nitrogen	Radius (N)/ Å	Calculated solv.energy	Experimental solv.energy	Deviation (ΔG)
Methylamine	1.75	-15.51	-19.10	3.59
Methylammonium	1.75	-304.01	-319.70	15.69
Pyridine	1.75	-23.55	-19.70	-3.85
Methylamine	1.70	-16.14	-19.10	2.96
Methylammonium	1.70	-304.37	-319.70	15.34
Pyridine	1.70	-24.61	-19.70	-4.91

Table 3-3. Calculated and experimental solvation energies with different radii

As can be seen (Table 3-3), a better fit to the experimental data was obtained by changing the chlorine radius to 2.15 Å. At this point, the deviation of the chloride anion is 0.4 kJ/mol, and -7.9 kJ/mol for chlorobenzene. The results for chlorobenzene are better when the radius of the chlorine is 2.2 Å but with this radius of the chloride anion the energies of the other compounds get worse, so compromised decision would

be to keep the radius of the chlorine at the value of 2.15 Å. The same was the situation with the nitrogen. The best results obtained- a radius of nitrogen- 1.70 Å.

The final reparametrisation of the atom radii to get close to the experimental solvation free energies are given in Table 3-4.

$R_{\text{solv}}/\text{\AA}$		
Atom	Default	Optimized
H	1.30	1.30
C	2.00	2.00
N	1.70	1.83
O	1.72	1.72
Cl	2.05	2.15

Table 3-4. Default and optimised solvation radii used in the calculations

3.2.3 Transition state search

For the purpose of the transition state searches, energy profiles with constraints, and restraints, with the possible chlorinating agents, were made. *TURBOMOLE*'s package was used for the constrained calculations with one bond distance fixed. The restrained calculations were done with *ChemShell*,^{89,90} with fixed one- and two-bond differences. Starting points for the transition state searches were the top stationary points of the restrained energy profiles. For the transition state searches *TURBOMOLE* provided energy and gradient for a microiterative algorithm in HDLCs (Hybrid Delocalized Internal Coordinates)⁹¹ and HDLCopt optimiser⁹¹ with low memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) optimiser^{92,93} for linear scaling geometry minimisation and partitioned rational function optimiser (P-RFO)^{94,95} for transition state searches implemented in *Chemshell*. After the transition states were found, numerical frequency calculations with the *NumForce*^{96,97} module with *TURBOMOLE*, were performed to prove the transition states. Another tool, which was used to verify the transition states, was the *TURBOMOLE*'s tool Screwer,

which distorts the transition state along to the imaginary vibrational mode. It was by this method that a search was done for the directly connected minima-products and reactants of our system.

3.3 Results and Discussion

3.3.1 Stabilities of chlorinated indoles

One of the main aspects of the project, in the beginning, was to perform calculations in order to investigate the stability of the tryptophan, chlorinated in different positions. Indole was used as a computational model (Figure 3-4). The type of reaction taking place is an electrophilic aromatic substitution, and minimisations of the σ -complex and product structures were conducted. The energies of the substituted chloroindole σ -complexes and products are put into a table and listed as a deviation from the most stable σ -complex and product (Table 3-5). The calculations showed that the most stable σ -complex is the σ -complex of 3-chloroindole. The second best is the σ -complex of 7-chloroindole. The most stable product is 7-chloroindole. In the enzyme 7-chlorotryptophan is formed because of its regioselectivity. The formation of 7-chlorotryptophan as product out of the enzyme confirms the theory that an enzyme does not change the intrinsic regiopreference of the reaction; it just lowers the energy barrier. Tryptophan chlorinating enzymes such as PyrH and Thal that do change the regioselectivity, catalysing the formation of 5 and 6 tryptophan substituted products.

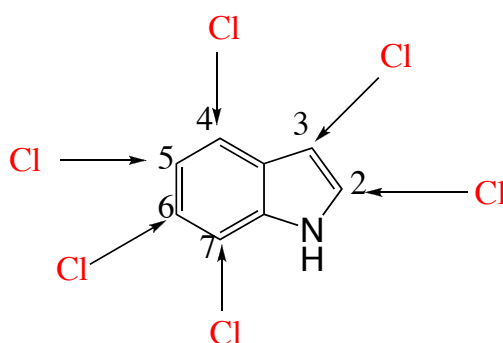


Figure 3-4 The investigated positions for the indole chlorination

	σ -complex	Product
Position of the substitution in the indole ring	Energy in kJ/mol	Energy in kJ/mol
C2	28.14	48.06
C3	-13.18	45.12
C4	29.98	33.07
C5	35.40	35.91
C6	26.70	35.86
C7	0	0

Table 3-5. Stability of the chlorinated compounds relative to 7-chloroindole (C7) at the TPSS/def2-TZVP+ level in IEFPCM water

3.3.2 Possible electrophiles involved in the indole chlorination

The possible reagents involved in electrophilic aromatic substitution in the system investigated for producing 7-chlorotryptophan were: 1) the free hypochlorous acid (HOCl); 2) the chloramine ($-\text{NH}_2\text{Cl}^+$) formed from the Lys78 residue of the enzyme and hypochlorous acid.

There are several possible reactions that could take place during the chlorinating process. Hydroperoxyflavin (FADHOOH) reacts with chloride to form the hypochlorous acid (Figure 3-5). The acid could stay free or interact with the Lys78 residue in the enzyme. It could form chloramine after a reaction between the acid and the enzyme residue Lys78 (Figure 3-6). The last possibility is the formation of a protonated hypochlorous acid (Figure 3-7)

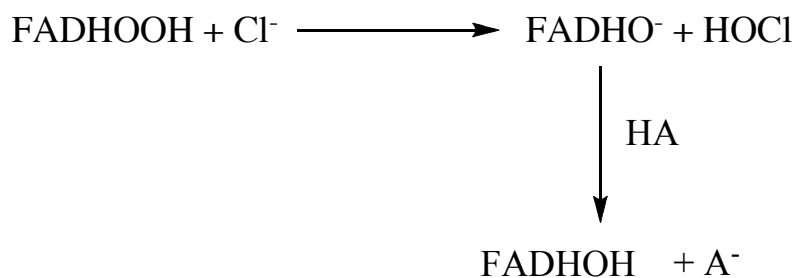


Figure 3-5. Formation of hypochlorous acid

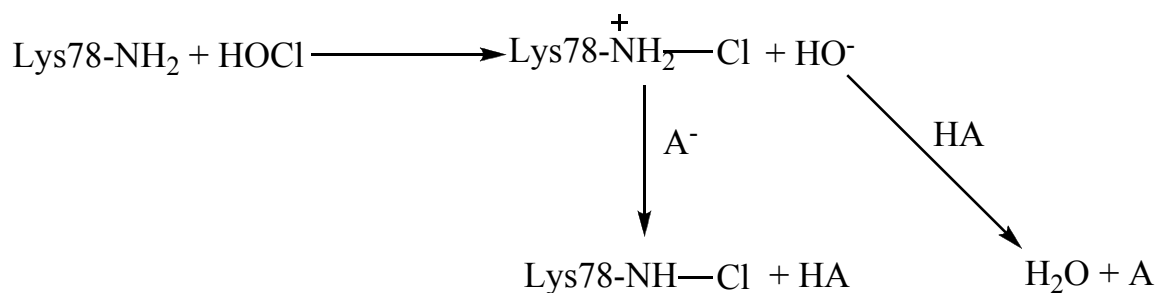


Figure 3-6. Reaction between hypochlorous acid and non-protonated Lys78

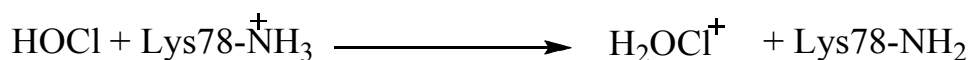


Figure 3-7. Reaction scheme for the formation of protonated hypochlorous acid

Calculations of forming the different chlorinating agents were made with our computational models:

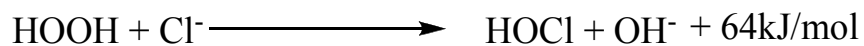


Figure 3-8. Formation of hypochlorous acid



Figure 3-9. Formation of a chloramine

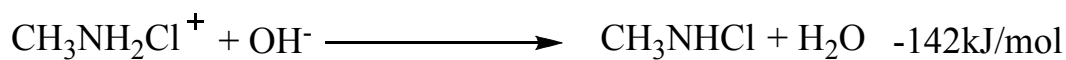


Figure 3-10. Formation of (chloroamino)methane



Figure 3-11. Formation of protonated hypochlorous acid

The central process in the mechanism of chlorination in PrnA is the formation of hypochlorous acid (Figure 3-8). (Chloroammonio)methane has an exothermic energy of formation (Figure 3-9). It is a product of a reaction between hypochlorous acid methylamine with reasonably high reaction energy. The reaction energy of the chloramine proves that it can be considered as a possible reagent, because of its possible existence in the solution. The chloramine could quite easily be an intermediate in a reaction with (chloroamino)methane as a final product (Figure 3-10). The protonated hypochlorous acid had a high energy of formation (Figure 3-11). It was found, after a literature search, that the existence of the protonated hypochlorous acid is under doubt. Previous calculations showed that the gas-phase proton affinity of HOCl is 153.1 kcal/mol at the CCSD(T)/6-311++G(3df, 3pd)//CCSD(T)/6-311(2df, 2p) level.⁹⁸ Spectroscopic results revealed that in low concentration chloride ion solutions, there are three species present: HOCl, OCl⁻ and Cl₂ (aq). Even at very low pH, the third species present was not H₂OCl⁺, as was expected to be.⁹⁹ In the solution was found only Cl₂ (aq) instead. From the results obtained, it was decided to rule out the protonated hypochlorous acid as being a possible halogenating agent.

3.3.3 Energy profiles for electrophilic chlorination of indole

The chosen reagents were (chloroammonio)methane + indole and hypochlorous acid + indole. The type of reaction is an electrophilic aromatic substitution. The first step represents the formation of a σ -complex and the second step involves a proton abstraction.

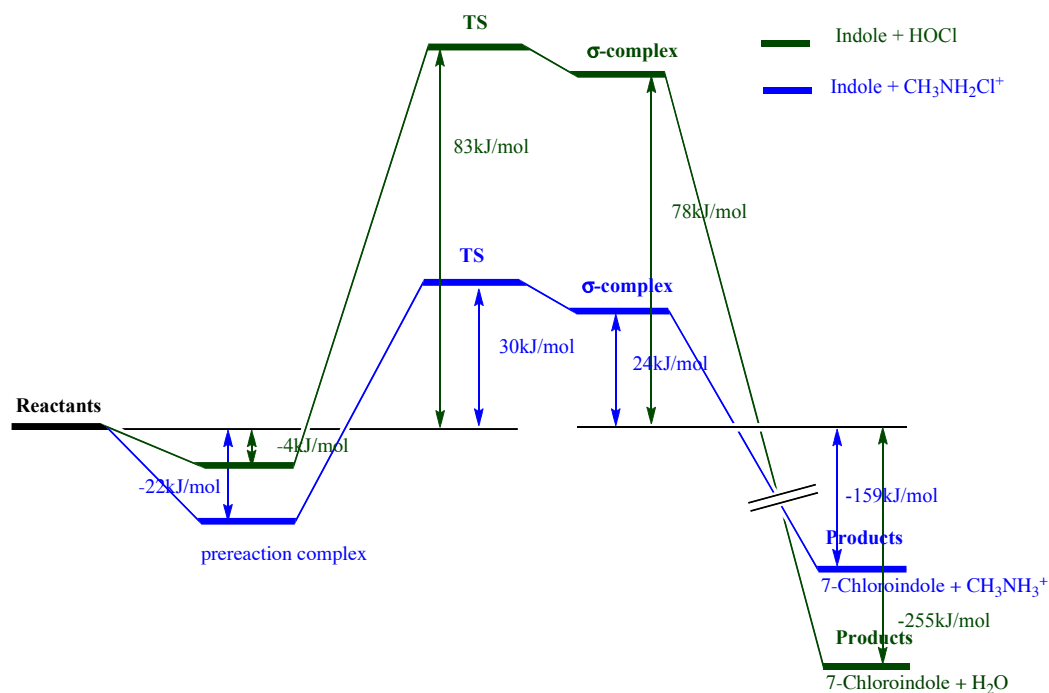
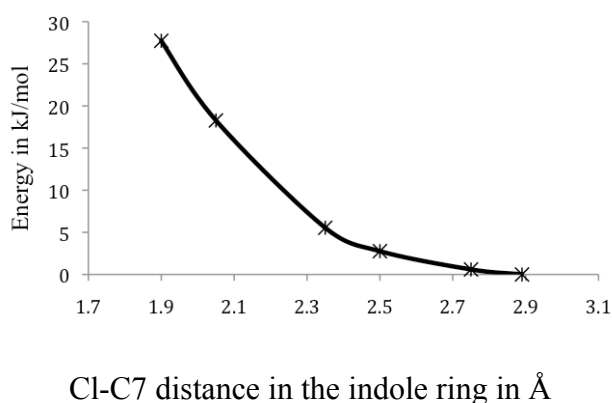


Figure 3-12. Potential energies of the σ -complexes of 7-chloroindole and products of hypochlorous acid + indole and (chloroammonio)methane + indole reactions, relative to the reactants at the TPSS/def2-TZVP+//IEFPCM($\epsilon_r = 78.4$) level of QM theory

A comparison of the energies of formation of the σ -complexes and the products of the chlorination of indole to 7-chloroindole with chlorinating agents- hypochlorous acid and (chloroammonio)methane was made. The transition states of the two proposed mechanisms were localised on the PES (Potential Energy Surface). The calculated energies of formation showed that the reaction energy ΔE_r of the σ -complex of 7-chloroindole with (chloroammonio)methane and indole as reactants was less endothermic (24 kJ/mol) than ΔE_r with hypochlorous acid and indole (78 kJ/mol) (Figure 3-12). The energy barriers ΔE^\ddagger of the 7-chloroindole σ -complex formation with hypochlorous acid and (chloroammonio)methane were 83 kJ/mol and 30 kJ/mol respectively. The formation of the σ -complex is the step that determines the energy of the overall reaction and would reveal the preference to either the (chloroammonio)methane or hypochlorous acid, as a chlorination agent. In this case, the formation of the σ -complex of (chloroammonio)methane is more favourable than the σ -complex of hypochlorous acid. The proton abstraction and formation of 7-chloroindole are strongly exothermic: -255 kJ/mol and -159 kJ/mol.

In order to obtain reaction profiles, constrained and restrained optimizations with *TURBOMOLE* and *ChemShell* were performed. The constrained calculations were along the C7-Cl bond distance (Figure 3-13). It was performed in order to locate a maximum along the chosen reaction coordinate. The first reaction that we looked at was the synthesis of the σ -complex of the pathway. Initially, in order to obtain the curve, the atom of chlorine was moved at different distances from the molecule of indole, freezing the distance between the atom of chlorine and the atom of carbon at position seven at the indole ring. Two curves were obtained, which showed the energies of the system as function of the distances between these atoms. Each of the points on the curves corresponded to the optimised geometries with the frozen Cl-C7 distances. That was done for both possible reactants and two plots were obtained.

A)



B)

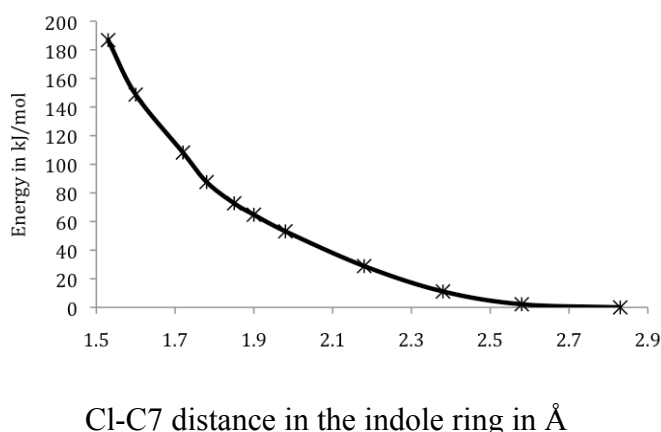


Figure 3-13. Energy profiles with optimised geometries at each point at the TPSS/def2-TZVP+ level with constraints of: A)(chloroammonio)methane + indole and B) hypochlorous acid + indole reactions

No maximum with constrained coordinates was found and for this reason restrained calculations were performed. The software used was *ChemShell*. The difference between the bonds of C7-Cl and Cl-O (N) in the reactions between hypochlorous acid and indole and (chloroammonio)methane and indole, was restrained. A harmonic restraint was applied. The force constant used was 30 Hartree/Bohr² (Figure 3-14).

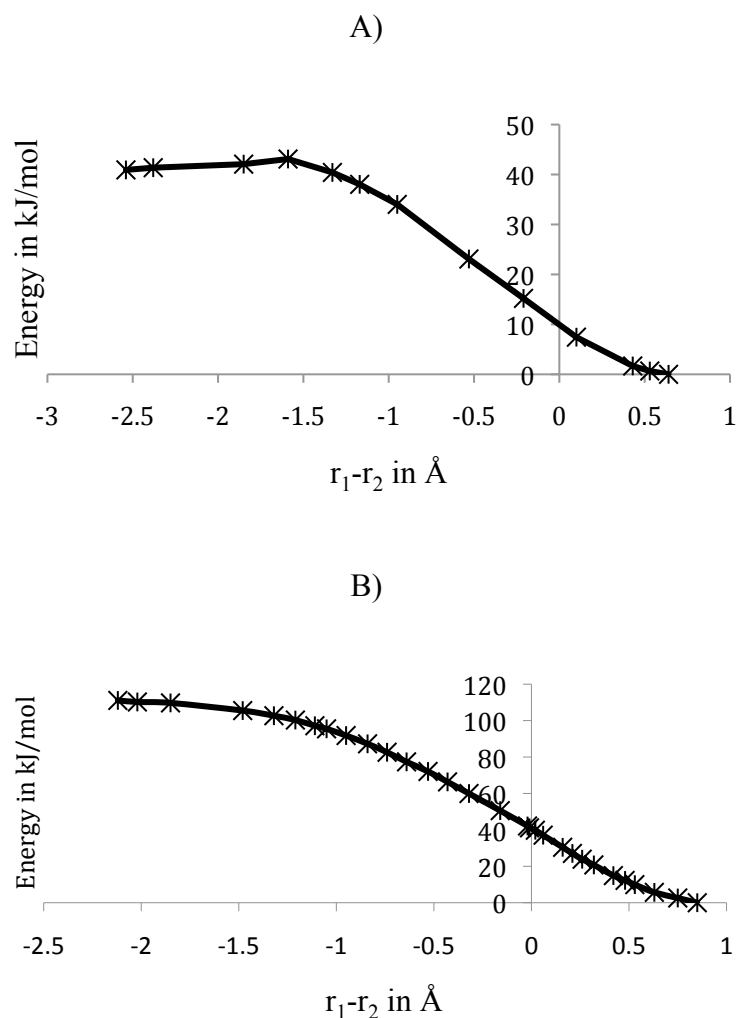


Figure 3-14. Energy profiles with optimised geometries at each point at the TPSS/def2-TZVP+ level with the restraints of: A) (chloroammonio)methane + indole and B) hypochlorous acid + indole reactions. The profiles start on the right side with the reactants and move along the curve in the direction of the products on the left side

Stationary points for both the (chloroammonio)methane + indole reaction and the hypochlorous acid + indole reaction were obtained as points at which the first derivative of the energy is equal to zero. These points were later used as starting points of the transition state searches.

3.3.4 Transition state search with *ChemShell*

The micro-iterative algorithm^{91,100} with *ChemShell* was used to perform a transition state search. The system was split into a reaction core containing very few atoms and an environment. While the environment was relaxed and the geometry was optimised using low memory Broden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm, the core was kept fixed. The optimisation of the environment was followed by one step at a time decoupled from the environment optimisation using the partitioned rational function optimiser (P-RFO) algorithm. The core atoms were the atoms forming the reaction core, in which a TS search was performed. The core atoms for hypochlorous acid+indole were 4, 11, 17, 18, 19 and for (chloro-ammonio) methane- 4, 11, 17, 18, 19, 20, 21, 22, 23, 24 (Figure 3-15).

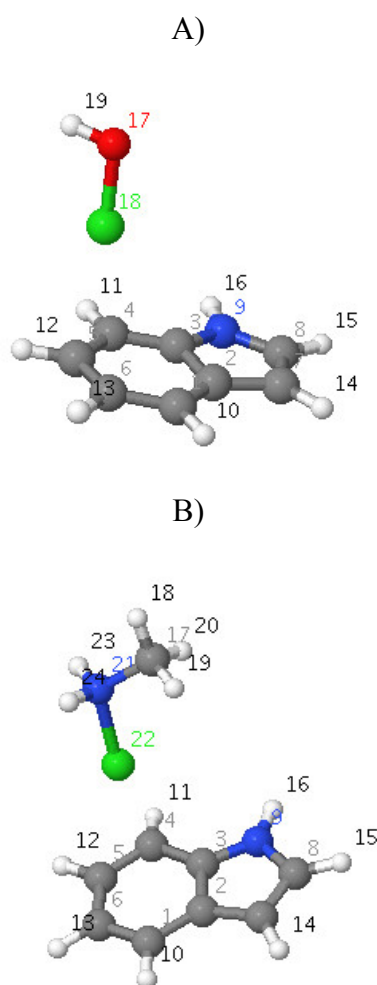


Figure 3-15. Starting structures and numbering of the atoms in the reactants: A) hypochlorous acid + indole; B) (chloroammonio)methane + indole

We successfully found the transition states of the hypochlorous acid+indole (Figure 3-16) and (chloroammonio)methane + indole (Figure 3-17) reactions.

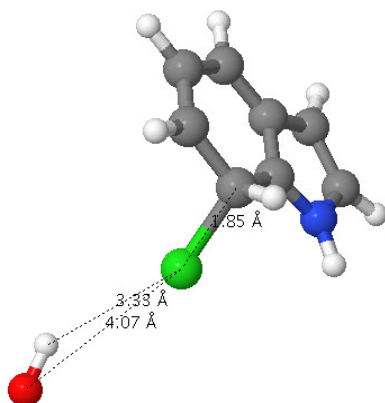


Figure 3-16. Transition state of hypochlorous acid + indole reaction at the TPSS/def2-TZVP+ level in IEFPCM continuum water

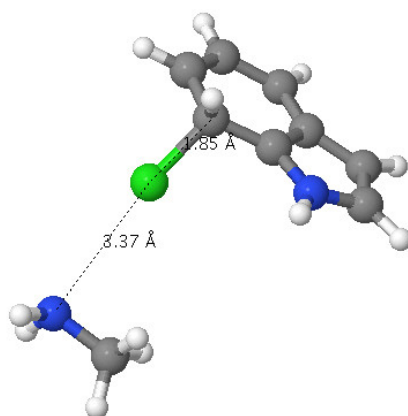


Figure 3-17. Transition state of (chloroammonio)methane + indole reaction at the TPSS/def2-TZVP+ level in IEFPCM continuum water

Compounds	$\text{CH}_3\text{NH}_2\text{Cl}^+$		HOCl	
Bond	C7-Cl	Cl-N	C7-Cl	Cl-O
Distance in Reactants	-	1.76 Å	-	1.61 Å
Distance in Transition state	1.85 Å	3.37 Å	1.85 Å	4.07 Å
Distance in Products	1.83 Å	-	1.83 Å	-

Table 3-6. Distances in reactants, transition states and products of the chlorinating process, with two different chlorinating agents

In the transition state structures, Cl is at a distance of 1.85 Å from C7 for both reagents (Figure 3-5). The transition state structures are close to the products, where this bond length is 1.83 Å. According to the Hammond's postulate, the transition state is expected to be close to the products for an endothermic reaction.¹⁰¹ This is in good agreement with our results. For the transition state with hypochlorous acid, the hydrogen of the hydroxide anion is orientated to the chlorine at a distance of 3.33 Å and this is not consistent with being a hydrogen bond present between the hydrogen of the hydroxide and the chlorine. In both transition state structures, the carbon C7 is pyramidalised and the hydrogen atom in the C7-H bond is no longer in the plane with the ring, as in the indole reactant. The Cl-O bond length of the hypochlorous acid in the reactant is 1.61 Å, and in the transition state that distance is 4.07 Å. The Cl-N bond in (chloroammonio)methane is 1.76 Å, and in the transition state this bond length becomes 3.37 Å.

The barrier, relative to the reactants for the hypochlorous acid and indole reaction, is 83 kJ/mol, and for the (chloroammonio)methane and indole reaction, it is 30 kJ/mol. There is a difference between the two barriers of 53 kJ/mol.

To confirm the transition states, *NumForce* calculations were performed with *TURBOMOLE*. The vibrational frequency calculations of the transition states of indole + hypochlorous acid and indole + (chloroammonio)methane yielded each, one imaginary frequency of -42.88 cm^{-1} and -49.90 cm^{-1} .

3.3.5 Pre-Reaction Complexes

The *TURBOMOLE* calculations performed on both reactants separately, and in one calculation, showed that between the reactants and the transition states there are stabilised structures present, which are pre-reaction complexes. The pre-reaction complexes have energies lower than the transition states and the reactants. The pre-reaction complex of (chloroammonio)methane is favoured over the reactants by 22 kJ/mol (Figure 3-18). The pre-reaction complex of the hypochlorous acid is 4 kJ/mol more favourable than the reactants (Figure 3-19).

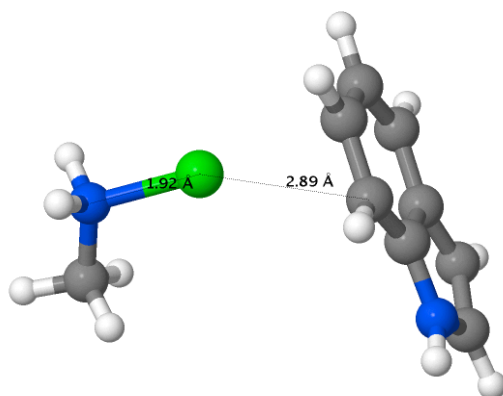


Figure 3-18. Pre-reaction complex of (chloroammonio)methane and indole at the TPSS/def2-TZVP+ level in IEFPCM continuum water

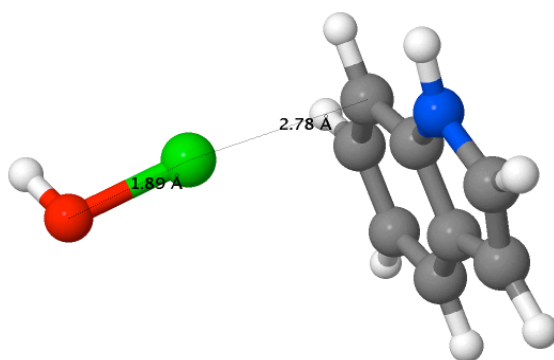


Figure 3-19. Pre-reaction complex of hypochlorous acid and indole at the TPSS/def2-TZVP+ level in IEFPCM continuum water

A comparison of the energies of the pre-reaction complexes of both chlorinating agents showed that the pre-reaction complex of (chloroammonio)methane, like the transition state, is of lower energy than the pre-reaction complex and transition state of the hypochlorous acid. The pre-reaction complexes proved yet again that the (chloroammonio)methane could be a better chlorinating agent than the hypochlorous acid.

3.4 Conclusions

The objectives for performing the calculations, presented in this chapter were to find whether the regioselectivity of forming 7-chlorotryptophan is intrinsic or enzyme-

induced and to investigate the reactions with possible chlorinating agents at the QM level as a preparation for QM/MM modelling.

A model system with indole for tryptophan and methylamine for the side chain of Lysine78 was used to investigate the chlorination of tryptophan. The calculations were performed at the TPSS/def2-TZVP+ level with QM-COSMO continuum model for water, using *TURBOMOLE*. The relative energies of all the monochloro-indole derivatives were calculated along with their cationic σ -complex intermediates. The most stable σ -complex was that of 3-chloroindole, and 7-chloroindole was the second best. The preferred product of the chlorination of indole was 7-chloroindole, proving that the intrinsic reactivity matched the product in the enzyme. The possible chlorinating agents H_2OCl^+ , HOCl and $\text{CH}_3\text{NH}_2\text{Cl}^+$, generating the “ Cl^+ ” for the electrophilic aromatic substitution in the indole ring were investigated. The protonated hypochlorous acid showed results of very high energy of formation and was ruled out as chlorinating agent. The hypochlorous acid was regarded to be formed after a nucleophilic attack of “ Cl^- ” on FADHOOH in the enzyme. If the acid reacts with the adjacent Lysine78 residue - the other viable chlorinating agents is formed: $\text{Lysine78NH}_2\text{Cl}^+$ ((chloroammonio)methane was used as a model).

The chlorination of indole with both chlorinating agents- HOCl and $\text{CH}_3\text{NH}_2\text{Cl}^+$ is an electrophilic aromatic substitution. The electrophile is “ Cl^+ ”. After the attack on the indole ring, a pre-reaction complex was formed, followed by the formations of the σ -complex of 7-chloroindole and the product 7-chloroindole. The reaction energies were strongly exothermic: -159 kJ/mol with HOCl as chlorinating agent and -255 kJ/mol with the electrophile $\text{CH}_3\text{NH}_2\text{Cl}^+$. The energy barriers and the reaction energies for the formation of the σ -complex of 7-chloroindole were: 30 kJ/mol and 24 kJ/mol for the (chloroammonio)methane + indole reaction, and 83 kJ/mol and 78 kJ/mol for the hypochlorous acid + indole reaction respectively. The (chloroammonio)methane is the favoured chlorinating agent over the hypochlorous acid but both are plausible. It means that probably the chlorinated lysine residue of the enzyme is a better chlorinating agent than the free hypochlorous acid.

As a future work other functionals could be tried and the results could be compared to those presented in this chapter. For our purpose the TPSS local functional was used because RI is implemented in *TURBOMOLE* only for pure functionals.

Chapter 4 Classical Molecular Dynamics simulations on PrnA monomer

4.1 Introduction

Initial preparation of the PDB (Protein Data Bank)⁶⁶ file of the halogenase PrnA (PDB ID 2AQJ) had to be done before any computational studies on the enzyme could be carried out. The work was accomplished by Christopher Hamilton.¹⁰² It included assigning the protonation state of the enzyme's residues by adding hydrogen atoms and deciding on the correct orientation of the side-chain of the amino acid residues. The programs Reduce,¹⁰³ What If,¹⁰⁴ PropKa,^{105,106} MCCE (Multi Conformational Continuum Electrostatics)¹⁰⁷ and H++¹⁰⁸ were used.

It is easy to predict the position of the unambiguous hydrogen (without any free degree of rotation). Polar hydrogen atoms are, on the other hand, quite problematic, especially if they have rotational degrees of freedom. These atoms' orientations usually depend on local hydrogen bond networks and steric effects in the structure.

The preparation of an enzyme structure passed through the following steps:

1. Validation of the PDB structure
2. Cleanup- removing all the artificial residues from the structure
3. Positioning of the unambiguous hydrogen atoms
4. Checking for flipped amide and imidazole chains- those chains that can be found in residues as Asn, Gln and His
5. Positioning of polar hydrogen atoms of Ser and Tyr
6. Positioning of ionisable atoms of His, Asp and Glu

All these steps were executed during the preparation of the PrnA structure. The data were compared after each step of the process. The titratable residues' protonation states were decided using the Henderson-Hasselbalch equation:

$$pH = pKa + \log \frac{[A^-]}{HA} \quad (4-1)$$

PrnA has a functional pH of 7.2.⁶⁴ From the equation above it can be concluded that any residue with pKa greater than pH+2 would be fully protonated and deprotonated if it has a pKa less than pH-2. That determined a range of pKa (5.2-9.2), which gives definite information about the protonation state of the amino acid- from definitely protonated to definitely non-protonated.

The most interesting results concerned the protonation states of the two residues involved in the possible mechanism of chlorination- Glu 345 and Lys78 (Table 4-1). The results suggested that Glu345 would be deprotonated, but the protonation state of the Lys78 remained unresolved.

		PropKa	MCCE	H++
Residue		pKa	pKa	pKa
GLU	345A	0.53	<0	-19.39
GLU	345B	0.89	2.01	-19.38
LYS	78A	5.48	8.30	16.94
LYS	78B	5.48	8.27	15.62

Table 4-1. pKa values for the important residues

4.2 Computational methodology

Classical Molecular Dynamics simulations on PrnA were performed with the GROMACS 3.3 program package,¹⁰⁹⁻¹¹¹ with the AMBER03 force field.¹¹² The results were analysed with the RMSD trajectory tool of VMD¹¹³ and plotted with the use of xmgrace and tcl script, specially written for the purpose.

Antechamber,¹¹⁴ an accessory module in the AMBER 10 package¹¹⁵⁻¹¹⁷ with the Leap tool¹¹⁴ was used for generation of GAFF (Generalised Amber Force Field)¹¹⁸ and ffAMBER03 residue topologies. The point-charges of FAD (flavin adenin dinucleotide), the zwitterion of tryptophan (ZTRP) and the peroxyflavin (FADHOOH) ligands were obtained by fitting to the electrostatic potentials of these compounds, calculated at B3LYP/cc-pVTZ//IEFPCM($\epsilon_r=4.24$)/HF/6-31G** level.

4.2.1 Residues building

The AMBER03 did not have residue topologies for the ligands present in the active centre of the enzyme- ZTRP (tryptophan substrate), FAD and FADHOOH. That necessitated for additional building of these residues. Calculations at the B3LYP/cc-pVTZ//IEFPCM($\epsilon_r=4.24$)/HF/6-31G** level of FAD, ZTRP and FADHOOH were made. This is the standard procedure for obtaining the MM charges in the AMBER03 force field established by Kollman and co-workers, who developed additionally the earlier approach of Cornell *et al.*¹¹² Constraints of torsions and angles with atom numbers: (39, 41, 42, 44); (41, 42, 44, 46); (42, 44, 46, 48); (44, 46, 48, 49); (46, 48, 49, 50); (48, 49, 50, 53); (49, 50, 53, 1); (50, 53, 1, 4); (53, 1, 4, 5); (1, 4, 5, 6); (4, 5, 6, 77); (42, 43, 58, 83); (78, 46, 47, 54); (46, 48, 49); (48, 49, 50) for FADHOOH and FAD were used for keeping the structures similar to X-ray structures and avoiding any further changes, caused by the optimisation (Figure 4-1). The single point calculations of ZTRP, FAD and FADHOOH were made in IEFPCM^{119,120}($\epsilon_r=4.24$). The choice $\epsilon_r=4$ by Kollman and co-workers reflected their intention to mimic the protein interior and to avoid overpolarisation. The overpolarisation would exaggerate the charges and the built residues would be too hydrophilic. The HF/6-31G** geometry optimisations of FAD and FADHOOH were performed in gas phase, but those of ZTRP were made in IEFPCM($\epsilon_r=4.24$). In the gas phase, the geometry optimisation of ZTRP resulted in a proton transfer occurring, with a proton being transferred from the ammonium group ($-\text{NH}_3^+$) to the carboxy group ($-\text{COO}^-$). Such a transfer did not occur in IEFPCM ($\epsilon_r=4.24$) (Figure 4-2).

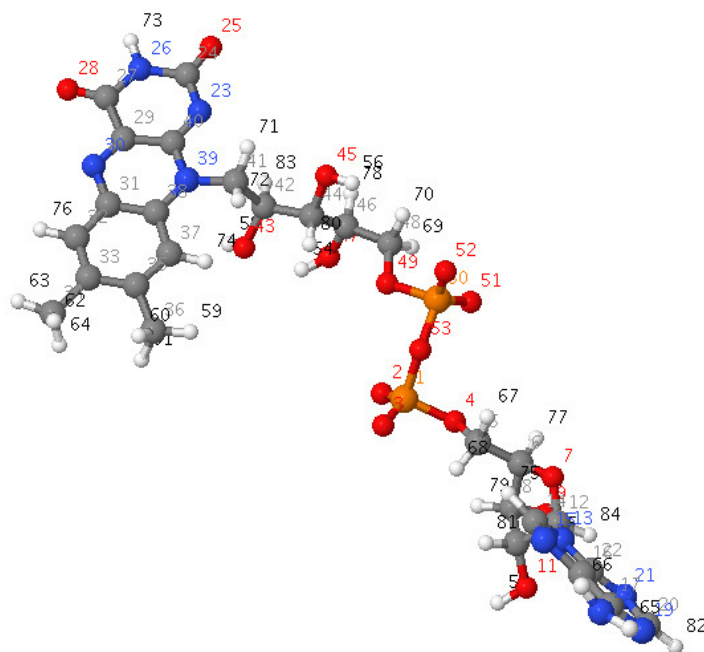


Figure 4-1. Optimised FAD structure at the HF/6-31G** level

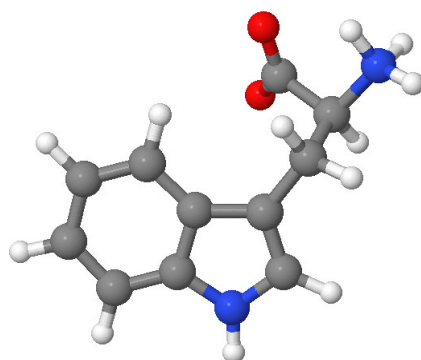


Figure 4-2. Optimised ZTRP structure at the HF/6-31G**//IEFPCM($\epsilon_r = 4.24$) level

The C-O-O-H torsion values associated with the minimum points of the scan with 5-methyl-4 α -peroxy-10-H-isoalloxazine at the B3LYP/cc-pVTZ level in gas phase, were used in the optimisation with the entire FADHOOH molecule. The optimisations with B3LYP/cc-pVTZ//IEFPCM($\epsilon_r = 4.24$)//HF/6-31G** started with 120° and -90° torsion orientation of the peroxy group in FADHOOH. Two possible minima at 105.1° and -78.96° were revealed (Figure 4-3 and Figure 4-4).

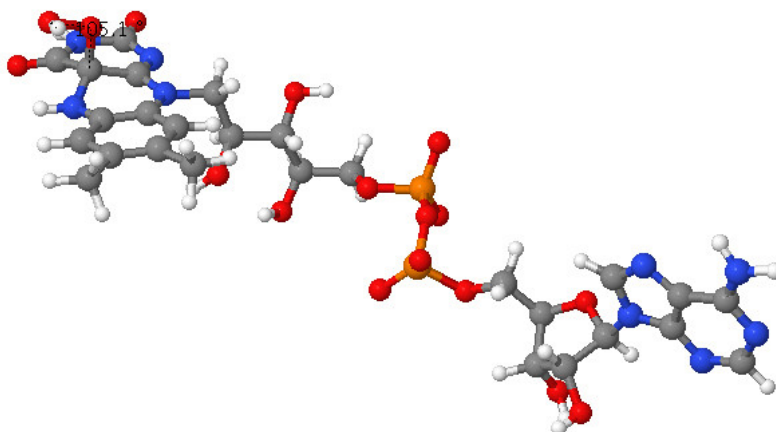


Figure 4-3. FADHOOH optimised structure at the HF/6-31G** level with 105.1° value of the C-O-O-H torsion

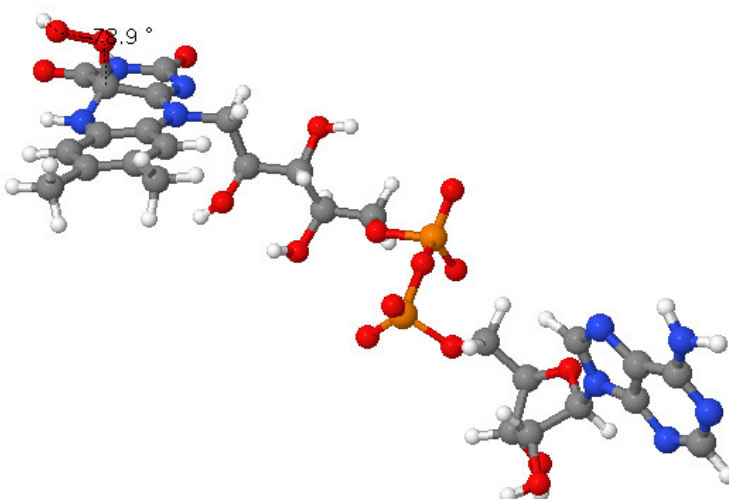


Figure 4-4. FADHOOH optimised structure at the HF/6-31G** level with -78.9° value of the C-O-O-H torsion

The optimised structures of ZTRP, FAD and FADHOOH were used for obtaining the *RESP* (restrained electrostatic potential) charges¹²¹ with the use of *Antechamber*. For FADHOOH the average charges of the two conformers were taken.

The topology of the three residues was built with the use of *Antechamber* and *XLeap*, both programs being part of the AMBER package. The files were converted from AMBER to GROMACS format with *amb2gmx* perl script, for obtaining the ffAMBER03 parameters in GROMACS. All the non-standard residues parameters are listed as follows:

4.3 Results and Discussion

4.3.1 MD simulations with FADHOOH in the PrnA monomer

4.3.1.1 Preparation of the initial structure for MD simulations and simulation parameters

As initial structure, the structure from the 2AQJ.pdb file was used⁶⁴ with assigned protonation states of the amino acid residues. Some of the atom names in the .pdb file were converted so they were compatible with the names of the same type of atoms in GROMACS with AMBER03 force field.

The first step in the protein preparation of the protein in GROMACS included the choice of the box. Different types of boxes were tested.

Box type	Cubic	Dodecahedron	Octahedron
Volume (nm ³)	1014.22	717.16	1014.22

Table 4-2. Volumes of box types for the PrnA monomer

The dodecahedral box was chosen with a minimal distance of 1nm between the protein and the wall of the box. The criterion was to choose the smallest box possible that can fit the protein for maximum computational speed and the dodecahedral box was the best choice.

MD simulations were performed for the neutral Lys78 (LYN) and lysine78-residue protonated PrnA (LYP). Both systems were solvated and sodium cations (Na⁺) were added to make them neutral. Exact information about the number of atoms, residues and molecules is given in Table 4-3.

Residue	LYN-system	LYP-system
Amino acids	517	517
FADHOOH	1	1
ZTRP	1	1
Cl ⁻	1	1
Na ⁺	8	7
Water molecules	20885	20886
Total number of atoms	70977	70980

Table 4-3. Description of LYN and LYP systems

After preparing the structure we performed the simulation, choosing the following parameters:

1. Position restraints were applied for both systems on the heavy atoms and released during the simulations. That prevented any drastic changes and deviation in the structure along the equilibration usually due to large solvent forces when the solvent is still not equilibrated. We started the simulations with $100000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ restraints for all the atoms and we released them consequently to the final production run to the values of $100 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ for the protein and the ZTRP and $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ for the FADHOOH and the chloride. All the hydrogen bonds were constrained.
2. A time step of $dt = 0.002 \text{ ps}$ was used.
3. The maximum step size we chose was 0.1 nm
4. The tolerance (convergence criteria for the energy minimisation) was $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$
5. Particle-mesh Ewald method (PME)^{122,123} was applied to represent the Coulomb interactions and the switch method to account for Lennard-Jones potential
6. The Coulomb and van der Waals cut-offs were 1 nm
7. We used Periodic Boundary Conditions (PBC)

Detailed description of the parameter used during the different steps of equilibration, the minimisation of the structure and the production run are given in Table 4-4.

Parameters	Minimization	Eq 1	Eq 2	Eq 3	Eq 4	Eq 5	Eq 6	Eq 7	Eq 8	Eq 9	Eq 10
Ensemble	-	NVT			NPT						
Time	-	20ps	100ps	100ps	500ps	100ps	100ps	100ps	50ps	1ns	5ns
Restraints FAD, Cl ⁻ in kJ mol ⁻¹ nm ⁻²	100 000	100 000									
Restraints Protein, ZTRP in kJ mol ⁻¹ nm ⁻²		100000				1000	100	100	100	100	100
Integrator	SD	MD									
nsteps	1000	10000	50000	50000	250000	50000	50000	50000	25000	500000	2500000
Thermostat	BT	BT	BT	BT	BT	BT	BT	BT	NH	NH	NH
T-coupling	-	P, NP									
τ_T (ps)	-	0.01, 0.01									
T_0 (bar)	-	298, 298									
Barostat	-	-	-	-	BB	BB	BB	BB	PR	PR	PR
τ_P (ps)	-	-	-	-	1	1	1	1	5	5	5
P_0 (bar)	-	-	-	-	1	1	1	1	1	1	1

SD=Steepest descent, Eq=Equilibration, BT= Berendsen Thermostat, PR= Parrinello-Rahman, NH= Nosé-Hoover, P,NP= Protein, Non-Protein

Table 4-4. Simulation parameters



The minimised structures of LYN and LYP looked very similar (Figure 4-6). At this stage of the simulations the only significant changes occurred in the chloride environment, such as the formation of the hydrogen bond between the peroxy group of FADHOOH and the chloride and the reorientation of the protein residues and waters, and they did not depend on the protonation state of the system.

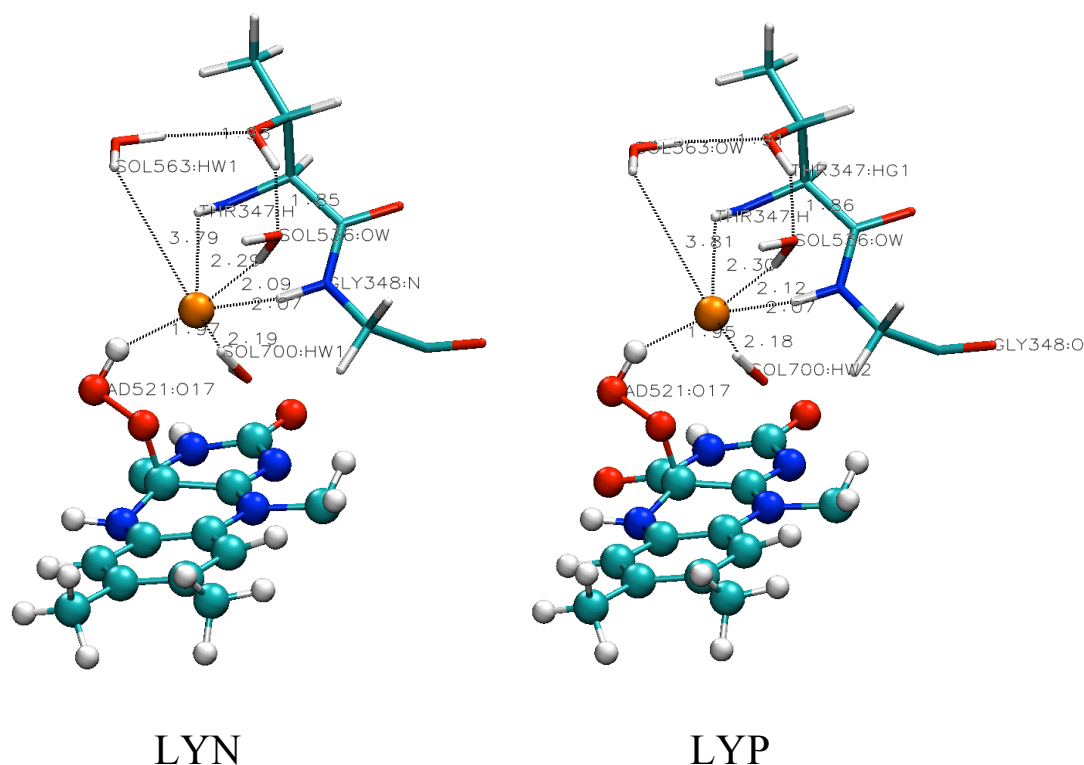


Figure 4-6. Minimised LYN and LYP structures with steepest descent using ffAMBER03. Hydrogen-bonding interactions of the chloride (orange)

Bigger differences can be expected in the Lys78 environment considering the difference in the protonation state of the lysine residue. In the PrnA initial structure Lys78 forms hydrogen bonds between, with its backbone to the backbone of Ala158. These hydrogen bonds change after the optimisation from 1.95 Å in the initial structure for H (Lys78)-O (Ala158) bond to 1.92 Å in LYN and 1.91 Å in LYP and from 1.64 Å for O (Lys78)-O (Ala158) bond to 1.70 Å in both LYP and LYP. In the minimised structures in both LYN and LYP hydrogen bonds of 2.26 Å and 2.04 Å with crystal waters are formed. The Lys78 in LYP forms a hydrogen bond with Thr50

of 2.03 Å in the initial structure and this hydrogen bond becomes shorter in the minimised LYP (1.97 Å) (Figure 4-7).

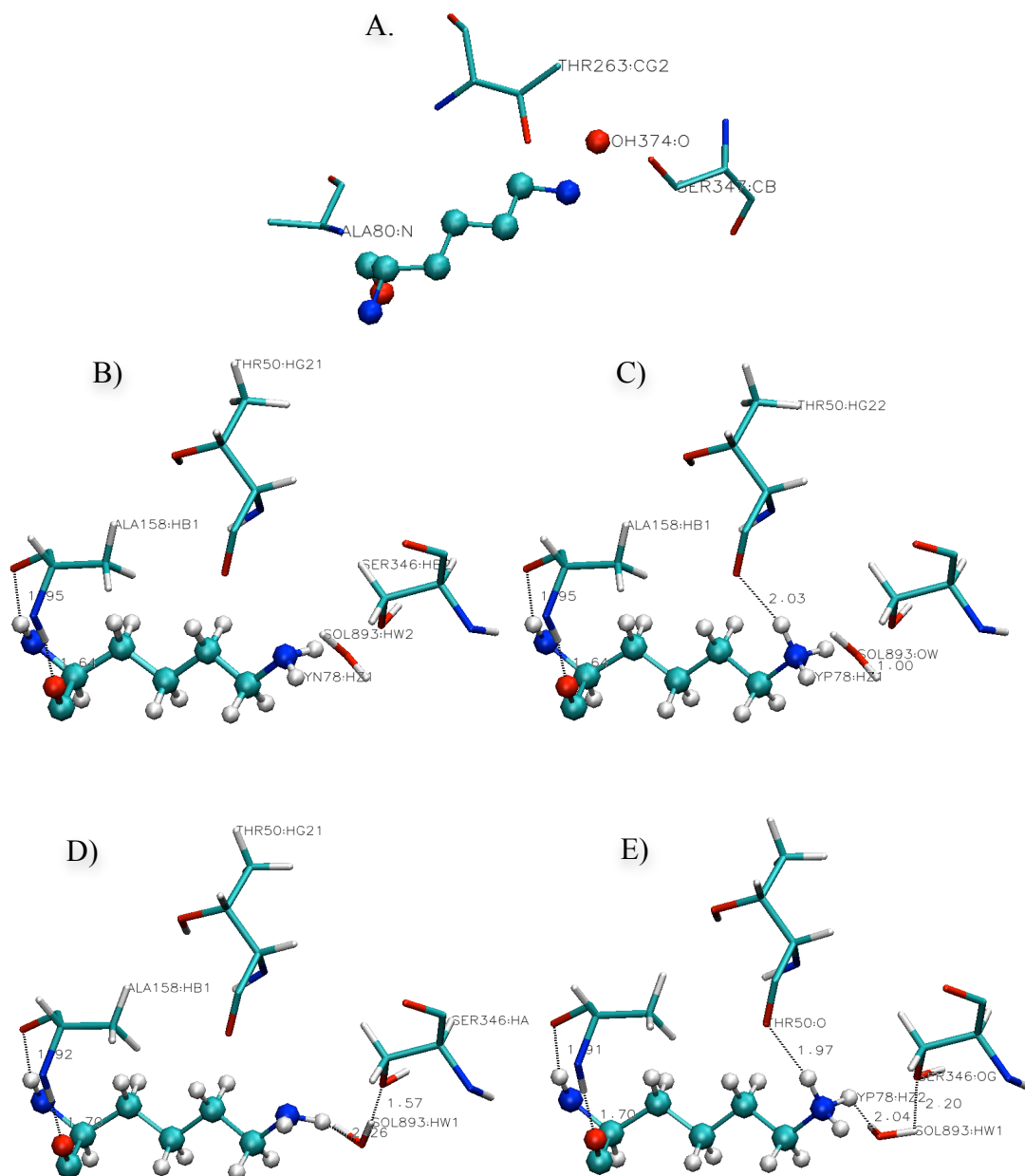


Figure 4-7. Lys78 structures: A) X-ray structure; B) and C) Initial LYN78 and LYP78 structures; D) and E) Minimised LYN78 and LYP78 structures using ffAMBER03

4.3.1.2 Total RMSD from X-ray structures

RMSDs (Root Mean Square Deviations) were to determine the deviation of the atoms of the simulated structures from the X-ray in ångströms (Å). The formula for calculating the deviation of a particular snapshot is as follows:

$$RMSD = \sqrt{\frac{\sum_{i=1}^N (r_{i,sim} - r_{i,expt})^2}{N}} \quad (4-2)$$

where N is the number of atoms in the structure and r represents the positions of the atoms in a Cartesian coordinate system with a centre of the mass at the origin of the simulated ($r_{i,sim}$) and experimental ($r_{i,expt}$) structures.

RMSD values from the X-ray structures for the entire LYN and LYP-systems at the last stage of the equilibration were calculated and compared (Figure 4-8). There were not drastic deviations of the total RMSD for both systems along the trajectory of this final 5ns of the simulations.

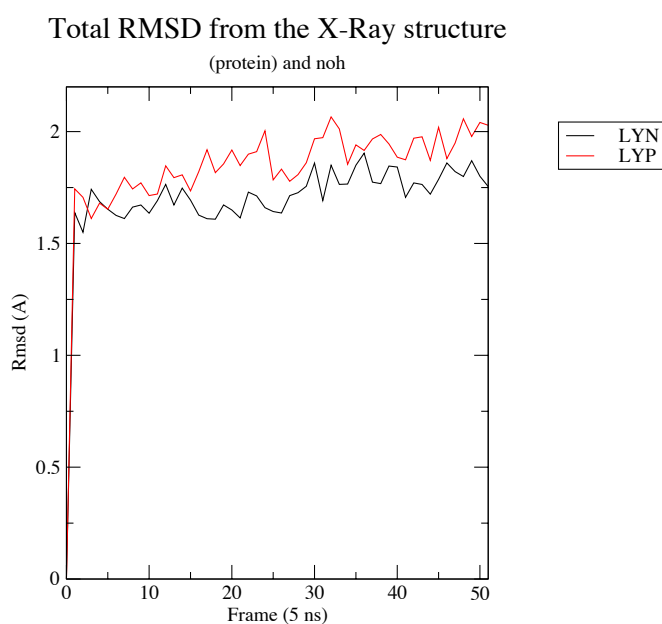


Figure 4-8. Total RMSD for all the protein atoms except the hydrogen from the production run

The average RMSD values for LYN and LYP systems aligned on the protein for the final 5 ns of simulation were 1.73 Å and 1.88 Å, respectively. The LYP system showed slightly higher deviation from the initial structure but, still, both systems stayed in the same range of structural resemblance- not more than 2 Å average

RMSDs. Looking at the deviations as a function of time, it was possible to conclude that both systems were stable and that they reached a point of equilibrium (Table 4-5).

Time	LYN	LYP
0-0.8 ns	1.65 Å	1.71 Å
0.8-1.6 ns	1.68 Å	1.78 Å
1.6-2.3 ns	1.66 Å	1.86 Å
2.3-3 ns	1.69 Å	1.89 Å
3-3.6 ns	1.79 Å	1.95 Å
3.6-4.3 ns	1.8 Å	1.94 Å
4.3-5 ns	1.8 Å	1.98 Å

Table 4-5. Deviation of all protein atoms except hydrogen of LYN and LYP systems from the X-ray structure, during the final 5 ns of the simulations

4.3.1.3 RMSD values of the protein residues

The calculated RMSD values per residue for this final 5ns of simulations showed that a couple of protein residues had an average deviation of more than 3 Å from the X-ray structure through the trajectory. The plotted data and the RMSD values are presented in Figure 4-9 and Table 4-6.

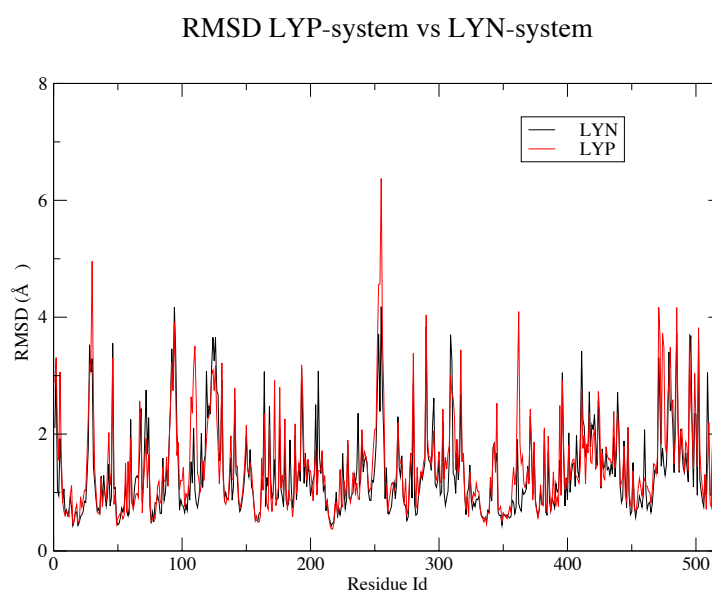


Figure 4-9. RMSD's averaged over the final 5ns of LYN and LYP protein heavy atoms from the X-ray structure per residue

Residue Id	Residue name	RMSD in LYN (Å)	RMSD in LYP (Å)
28	GLN	3.53	2.98
29	GLN	3.06	3.21
30	GLN	3.29	4.96
46	VAL	3.55	3.30
94	ASP	4.17	3.92
110	GLY	1.04	3.51
124	GLY	3.66	3.09
126	GLN	3.66	3.17
252	ASP	2.76	3.51
253	ASP	3.71	4.56
254	ALA	2.39	4.57
255	ARG	4.18	6.37
256	ASP	2.63	3.59
290	HIE	3.84	4.04
309	ASP	3.70	3.00
362	HID	0.81	4.10
471	ASP	3.31	4.17
472	ARG	2.46	3.67
474	LEU	1.75	3.73
485	GLU	3.73	4.17
495	ARG	3.32	3.71
496	ARG	3.67	2.49
502	ARG	3.59	3.82
516	ARG	6.39	6.51
517	ASP	4.72	7.57

Table 4-6. RMSD values per protein residue for the protein atoms except hydrogen averaged over the final 5 ns of the simulations

1. Residues 28, 29, 30, 110, 362 and 471 to 474, which all had high RMSD values associated with them, were present in a turn at the interface to the (omitted) second chain of the dimer.
2. Residues 94, 124, 125, 252 to 256 and 290 are in a turn of short helical fragment, respectively, exposed to solvent.

3. Residue 309 is placed on the surface of the protein.
4. Residues 482, 495, 496 and 502 are at the head of a helix, which is in contact with the residues 28-30 at the dimer interface.
5. Residues 516 and 517 are part of the C-terminal of the protein

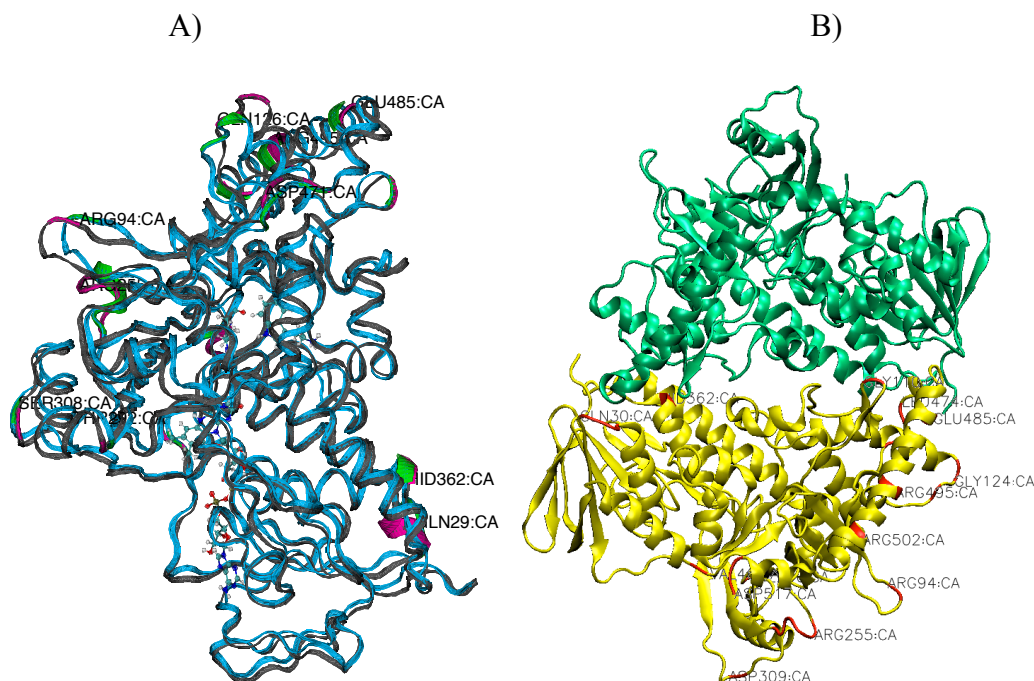


Figure 4-10. Positions of protein residues with high RMSDs. A) Deviation of the last snapshot of the simulated PrnA monomer (cyan) from the X-ray-structure (gray). The deviated residues are shown in green for the simulated PrnA, and in magenta for the crystal structure; B) PrnA dimer. The deviated residues are shown in red

4.3.1.4 RMSD values of the ligands

The RMSD values associated with the studied ligands stayed small and they almost stayed in the same position as in the crystal structure. They were again higher for the LYP-system compared to the LYN-system but still less than 2 Å. A more fair comparison here would be to the minimised structure than to the X-ray structure because the peroxy group of the FADHOOH was built by hand, especially for FADHOOH. The influence of the peroxy group is obvious for the deviation of the chloride in the LYN and the FADHOOH in LYP (Table 4-7).

Residue name	RMSD in LYN (Å) from the X-ray	RMSD in LYN (Å) after the minimisation	RMSD in LYP (Å) from the X-ray	RMSD in LYP (Å) after the minimisation
Cl ⁻	1.53	1.23	1.39	1.39
ZTRP	1.14	1.15	1.44	1.10
FADHOOH	0.90	0.89	0.92	0.77

Table 4-7. RMSD values of the ligands over the final run trajectory

4.3.1.5 Deviation in the Lys78 residue

The lysine78 residue showed practically no deviation in both systems after the simulations.

Residue name	RMSD in LYN (Å)	RMSD in LYP (Å)
LYS78	0.50	0.61

Table 4-8. RMSD values of the lysine 78 residue in LYP and LYN from the X-ray structure

4.3.1.6 The chloride binding site

4.3.1.6.1 LYN

The simulated LYN structure and the structure after the initial minimisation looked very much alike (Figure 4-11). There were small rearrangements of the hydrogen bonds and orientation of the water molecules and surrounding residues. Around 2.5ns of the simulations the distance between Wat700 and the chloride elongated. During the last nanosecond, Wat536 and Wat700 swapped their positions compared to what is seen in the minimised structure. That can be seen in the mirror image of their O-Cl⁻ distances along the simulation (Figure 4-12) and almost the same average value of those distances (Figure 4-17). In reality, the water molecules are indistinguishable, so it can be concluded that at the end of the simulation the structure looked almost pretty much the same regarding the position of the water molecules compared to the initial

minimised structure. The plot of the distribution of the distances between the chloride and the water molecules in the active site is presented in Figure 4-13. The distribution is larger, in the region of 3-3.5 Å. This confirms the existence of hydrogen bonds between the chloride and these residues.

The residue rearrangement around the chloride is consistent with the difference in the rotation of hydroxygroup of Thr347, which formed a hydrogen bond with the chloride instead of interacting with the closest water molecules. This happened along with the reorientation of Wat536 and Wat700 in the way that they formed a hydrogen bond between them.

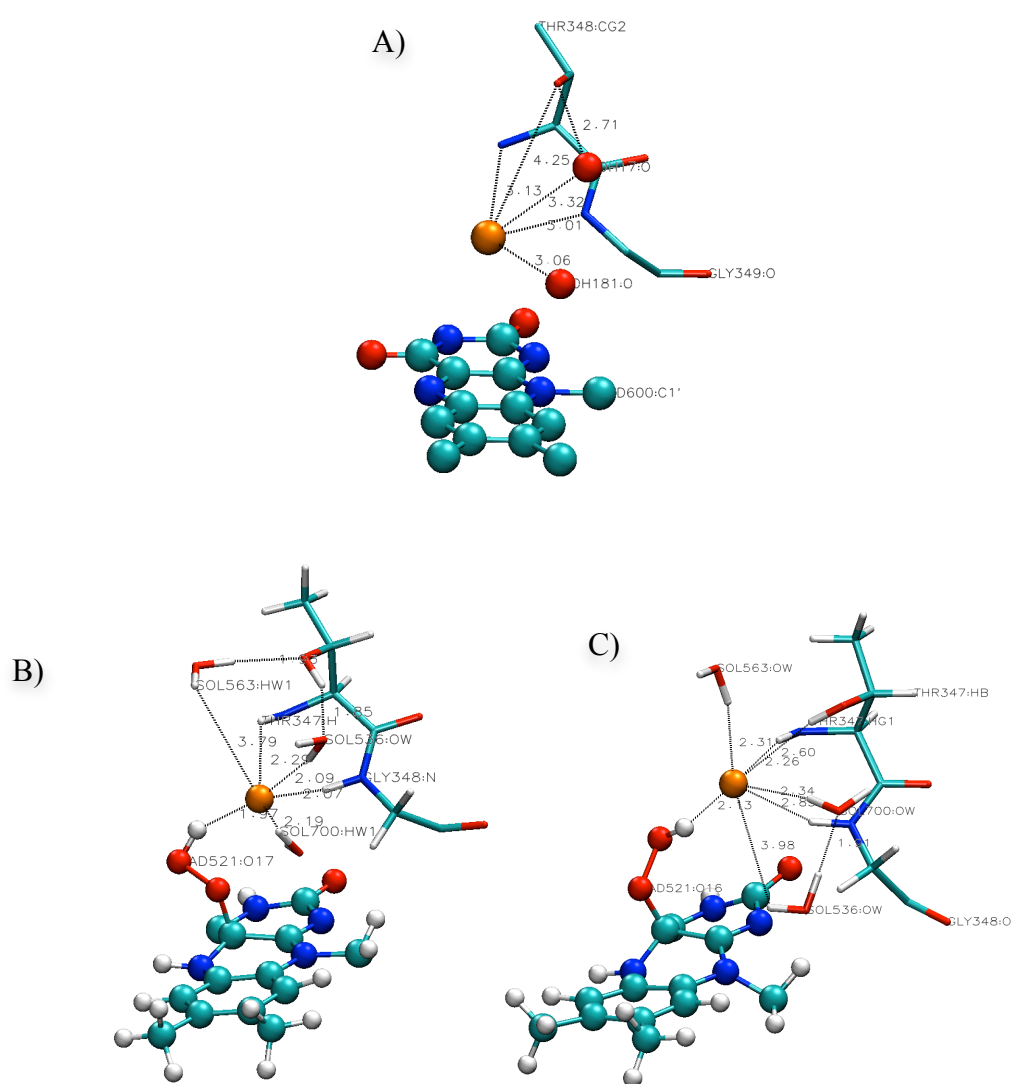


Figure 4-11. Chloride surroundings in the X-ray structure (A), after the initial minimisation (B) and after the production run of the MD simulation with LYN- last snapshot (C)

Distance between Cl519 and Wat 536, Wat563 and Wat700, in LYN

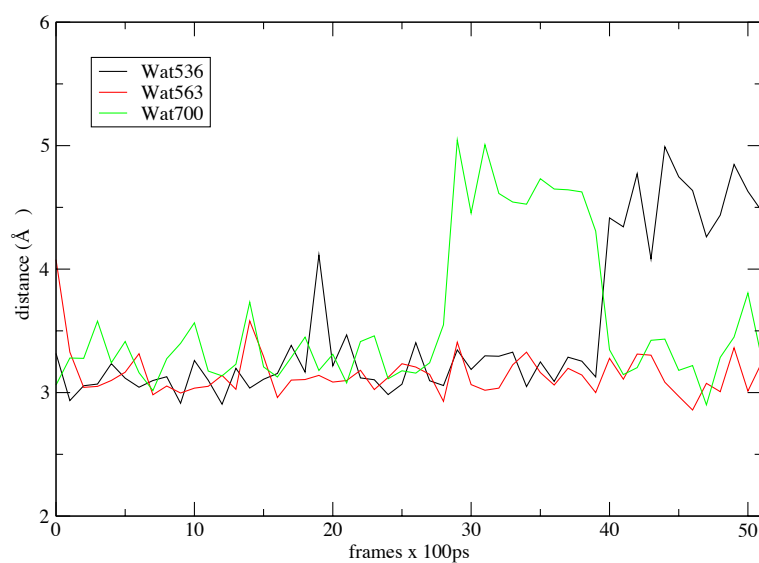


Figure 4-12. Chloride-water distances in the chloride binding site of LYN78 in the last 5ns of the simulation

Distribution of distances between C519 and Wat536, Wat563 and Wat700, in LYN

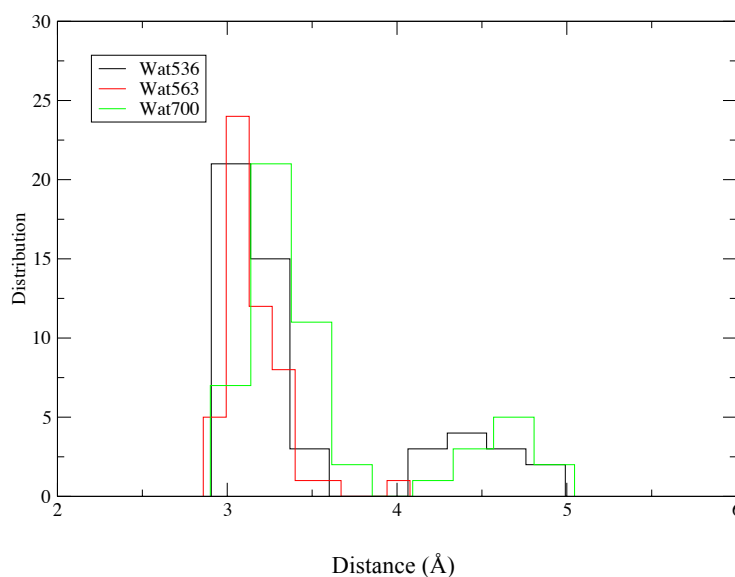


Figure 4-13. Distribution of the distances between the chloride and the oxygen atoms of water molecules 536, 563 and 700 in the LYN system

4.3.1.6.2 LYP

In the simulated LYP-system all the hydrogen bonds between the Cl^- -surrounding residues broke, in favour of formation of hydrogen bonds with the chloride itself

Distance between Cl519 and Wat536, Wat563 and Wat700, in LYP

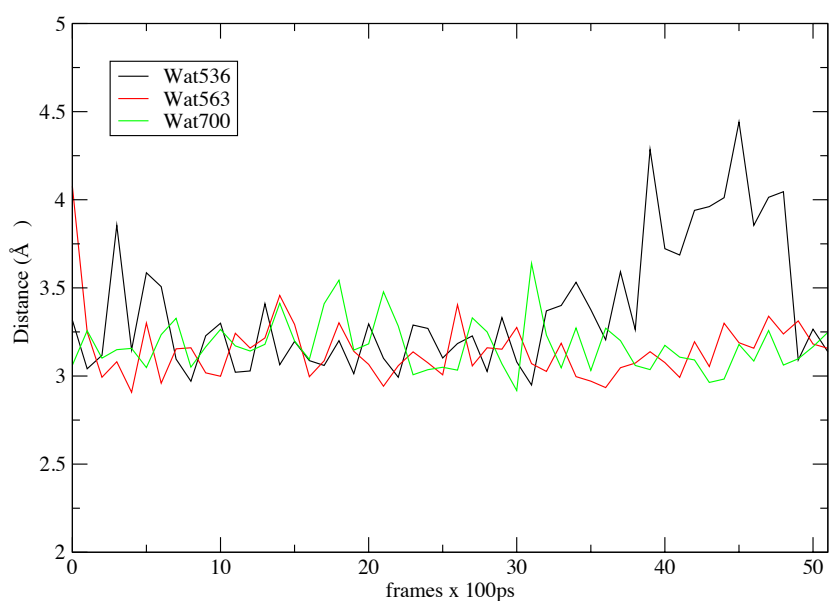


Figure 4-15. Chloride-waters distances in the chloride binding site of LYP78, in the last 5ns of the simulation

Distribution of distances between C519 and Wat536, Wat563 and Wat700, in LYP

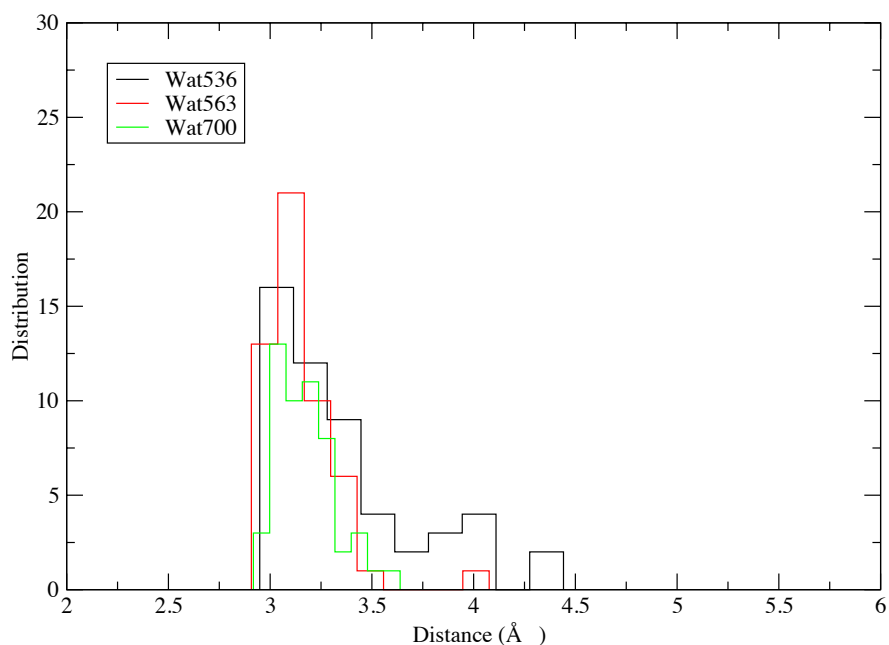


Figure 4-16. Distribution of the distances between the chloride and the oxygen atoms of water molecules 536, 563 and 700 in the LYP system

those of the FADHOOH simulations, except for the additional big block of residues 435-444. These residues are located in a helix, where the monomers are in contact with each other.

The RMSD values of the residues in the dimer were smaller in comparison to the RMSD values of the monomer. All the residues with high deviations from the X-ray structure were stabilised in the dimer (Table 4-9).

Residue Id	Residue name	Monomer RMSD in Å		LYP-Dimer RMSD in Å		LYN-Dimer RMSD in Å	
		LYP	LYN	Chain-A	Chain-B	Chain-A	Chain-B
28	GLN	5.97	4.70	2.30	2.23	2.04	1.80
29	GLN	5.7	2.88	2.40	2.32	2.36	2.00
30	GLN	6.02	2.39	2.74	2.11	1.92	1.79
125	PHE	6.18	2.49	2.48	1.34	1.76	1.31
254	ALA	5.06	1.34	3.53	2.34	3.41	1.99
255	ARG	6.82	2.34	3.91	4.02	5.61	3.03
309	ASP	4.89	2.06	3.13	2.50	4.66	2.79
435	THR	4.66	4.64	2.60	1.24	1.00	1.18
436	SER	10.01	9.14	2.98	1.50	1.55	1.83
437	PHE	8.66	7.74	3.16	1.73	1.73	1.68
438	ASP	9.69	8.54	3.27	1.96	1.87	1.86
439	ASP	9.26	7.46	3.37	1.88	3.89	3.53
440	SER	8.15	7.23	2.60	2.24	2.06	2.06
441	THR	5.68	5.50	2.13	0.81	1.37	1.02
443	TYR	5.88	5.69	3.06	1.24	1.17	1.06
444	GLU	4.77	5.20	3.06	1.65	1.89	1.74
482	GLU	3.76	5.50	2.73	1.70	1.97	2.34
485	GLU	5.4	3.78	3.44	3.27	2.56	3.01
516	ARG	4.76	4.70	8.38	6.47	7.46	6.99
517	ASP	5.84	1.93	3.45	2.58	6.15	6.79

Table 4-9. Average RMSD values of the residues present in the dimer interface in various molecules. (Chain A: Residues 1 to 521 Chain B: Residues 522 to 1042)

The results for the ligands showed deviation in the position of the chloride from the X-ray structure in LYN-system for both monomer and dimer. In LYP, the chloride

deviated less in chain A in the dimer than it did in the monomer and in chain B of the dimer. The stabilised residues 434-444 in the dimer kept the position of ZTRP. FAD stayed unchanged (Table 4-10).

Residue (ID)	Monomer		Dimer	
	RMSD in Å		RMSD in Å	
	LYP	LYN	LYP	LYN
Chloride (519)	1.24	4.31	0.90	5.13
ZTRP (520)	2.20	2.68	1.24	0.89
FAD (521)	0.80	0.98	1.03	0.89
Chloride (1040)	-----	-----	3.82	5.01
ZTRP (1041)	-----	-----	0.91	0.76
FAD (1042)	-----	-----	1.16	1.19

Table 4-10. RMSD values of ligands in monomer and dimer system

The last snapshot from the MD simulations with the LYP dimer with FAD was chosen for an initial structure in the QM/MM PrnA dimer simulations with FADHOOH. Both LYN and LYP structures were equilibrated by Sarath Chandra Dantu for 6ns. The chloride anion deviated less from the X-ray in chain A in the LYP dimer structure (0.90 Å) than in the LYN system (5.13 Å in chain A; 5.03 Å in chain B).

4.4 Conclusions

This chapter presents the results of the MD simulations on the PrnA monomer and dimer with FAD and FADOOH in water using ffAMBER03 under PBC conditions. The purpose was an equilibration of the solvated protein. The preparation of the enzyme included the choice of the protonated states of the protein's residues, the flips of the Asn and Gln residues and clarification of the protonation states of the Lys78 and Glu345 residues. Parameters for the non-standard residues (FADHOOH and

ZTRP) were derived using *XLeap* for the topology and *Antechamber* for the *RESP* charges.

After steepest descent minimisation and classical dynamics simulations with the FADHOOH ligand, the chloride in both LYN (PrnA with non-protonated lysine78 residue) and LYP (PrnA with protonated lysine78 residue) -systems formed a hydrogen bond with the peroxy group of FADHOOH. The existing hydrogen bonds with the surrounding water molecules and residues in the active site, stayed relatively stable.

The MD simulations on the PrnA monomer with bound FAD showed deviation of protein residues on the surface where the two chains interfaced with each other and that caused disturbance of the tryptophan-binding site. As a result the substrate ZTRP moved at 2.20 Å and 2.68 Å in LYP and LYN from its position in the crystal structure. When the simulations were conducted with the dimer, no significant alterations of the ZTRP occurred, in LYP its average RMSD values were 1.24 Å and 0.91 Å in chain A and B, and in LYN- 0.89 Å and 0.76 Å respectively. These small ZTRP's RMSDs in the dimer proved the need for the dimer for the future simulations. The last snapshot of the MD simulations on the LYP dimer with FAD was used as starting structure for the QM/MM modelling, in which the deviation of chloride anion in chain A (Cl519) was 0.90 Å, compared to LYN where the chloride (Cl519) moved at 5.13 Å from the X-ray. In chain B in both LYP and LYN, the chloride moved at great distance.

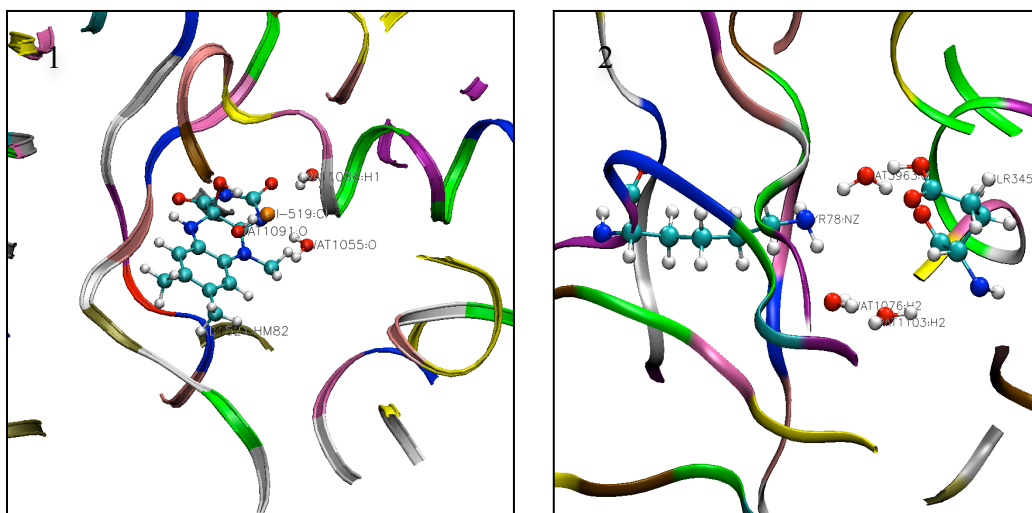
The MD simulations with LYP and LYN led to the conclusion that the protonation state of lysine78 residue has no significant effect on the structure.

As it was already mentioned before, the last snapshot of the equilibrated LYP dimer FAD was used for the QM/MM calculations. The FADHOOH ligand was built after adding a peroxy group to the FAD ligand. More snapshots along the trajectory of the simulations could be used as starting structures for the QM/MM modelling. Then pathways with different starting structures could be compared.

Chapter 5 QM/MM modelling on PrnA dimer

5.1 Introduction

QM/MM modelling is an approach for investigating reactions in big systems such as proteins and all the other biopolymers. In a system such as the PrnA dimer, thousands of atoms are involved and that raises the necessity for a combined method that describes the ongoing reactions and mechanisms. The active centre in PrnA, where the reactions take place, was described using a QM method. The QM method allowed to look at the bond breaking and bond forming processes. Depending on the step in the mechanism, different QM regions were chosen. The rest of the protein was described with MM parameters. The reactions of 1) the attack of the chloride on FADHOOH, 2) the formation of the chloroamine residue from Lyn78 and the hypochlorous acid (HOCl), 3) the exchange of a proton between the Lyp78 and Glu345 and 4) the step involving the chlorination of the substrate tryptophan, were the steps that were modelled in the QM/MM. Figure 5-1 presents the majority of the QM regions of these reactions:



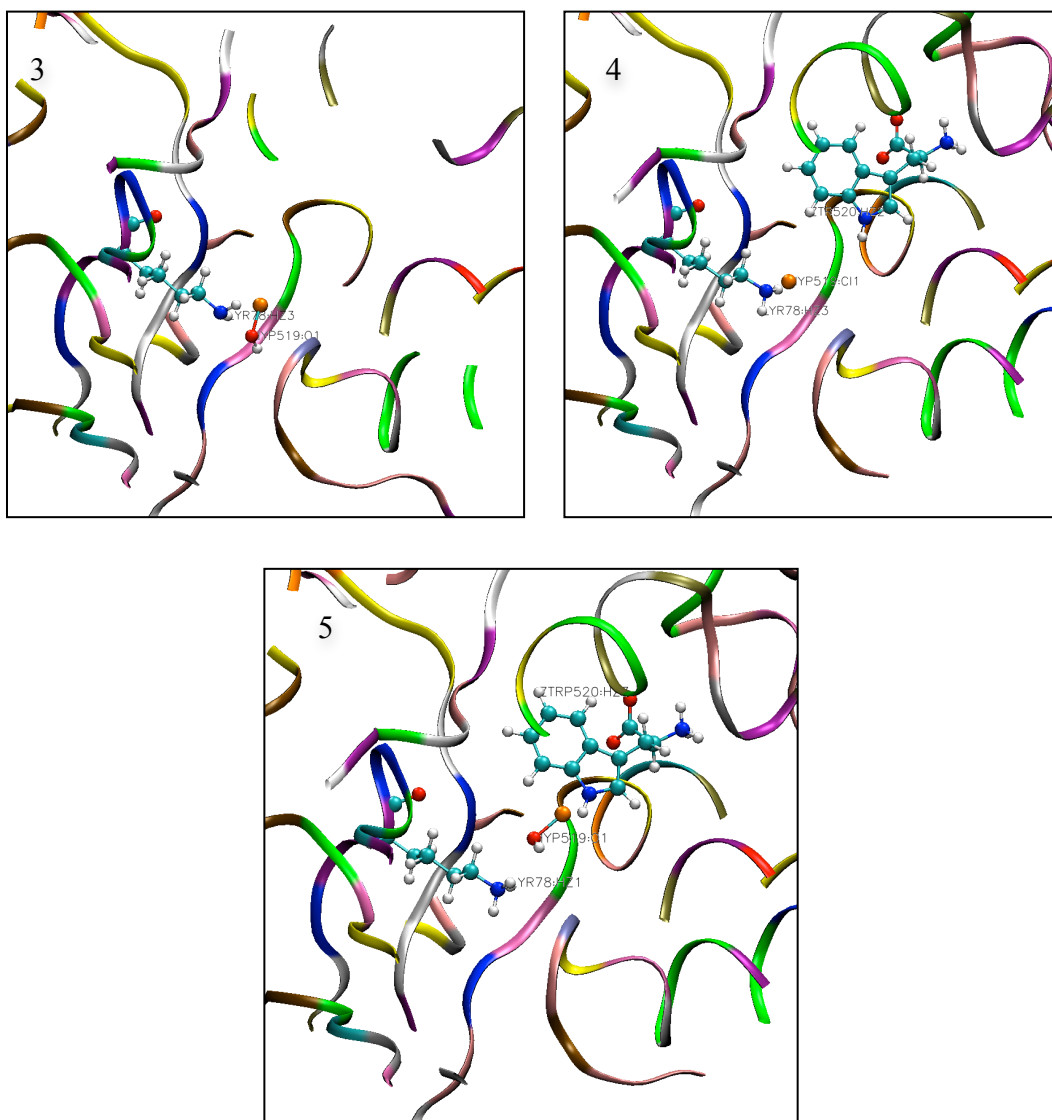
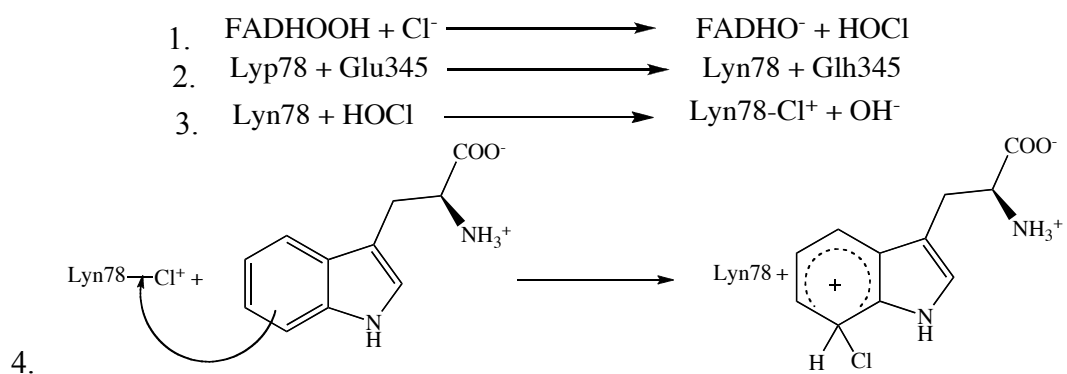


Figure 5-1. QM regions in the optimised initial structures at the AM1/AMBER03 level along the mechanism of the chlorination in PrnA

The main net reactions at Figure 5-1 can be presented as follows:



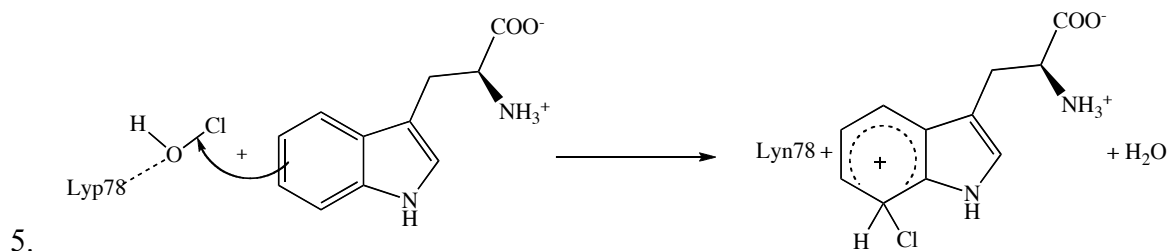


Figure 5-2. Mechanisms of the reactions, subject to QM/MM studies

In Figure 5-2 Lyn and Lyp represent non-protonated and protonated lysine residues. Glu and Glh are glutamate and glutamic acid.

5.2 Computational methodology

A preparation for QM/MM studies was carried out. It included a validation of the semi-empirical method that was to be used as a quantum mechanical method in the QM/MM simulations. A DFT B3LYP/6-31+G** scan around the C-O-O-H torsion in the peroxyflavin ring, using the *Gaussian 03*¹²⁴ computational package, and AM1²⁰ and PM3^{125,126} scans, using the *ChemShell*/MNDO interface, were performed. The aim of these calculations was to validate the semi-empirical method which will be used in the QM/MM modelling.

The QM/MM calculations were performed initially at the AM1/AMBER03¹¹² level and refined at DFT(TPSS)/AMBER03 level afterwards, using the *ChemShell* program package.^{89,90} The QM gradient and energy for AM1 were provided by the MNDO interface for semi-empirical calculations, and for TPSS the *TURBOMOLE* package⁷⁹⁻⁸³ was used. The TPSS functional was chosen as a non-local functional which permitted the use of the resolution of the identity (RI) approximation in *TURBOMOLE*. Def2-TZVP+ and def2-SVP+ augmented basis sets were applied in the DFT calculations. The MM region energy and gradient were calculated with the AMBER03 force field, using the DL_POLY molecular simulation package.¹²⁷ Electrostatic embedding described the interaction between the QM electron density and the MM point charges by including them in the QM Hamiltonian. Charge shift coupling was used at the boundary between the two regions without any QM/MM electrostatic cut-offs. At the QM/MM boundary the link atom scheme was used and

an H atom was added to the free valence cut bonds. The missing topologies for the tryptophan (ZTRP), peroxyflavin (FADHOOH) and hydroxyflavin (FADHOH), were built with *Xleap*¹¹⁴ and GAFF (Generalized AMBER Force Field).¹¹⁸ Their electrostatic potentials (*RESP* charges) were calculated at B3LYP/cc-pVTZ/6-31G** level and were applied in the QM/MM calculations using AMBER's additional tool- *Antechamber*. The normal optimisations were performed in HDLC coordinates along with the HDLCopt optimiser and L-BFGS algorithm. TS were found with the P-RFO algorithm. The explicit water molecules had O-H and H-H bonds constrained. The charges in residues where the C-C bonds were cut and the residues were in both QM and MM regions were redistributed within the residue to add to an integer number for both regions.

5.2.1 Validation of O-O torsional parameters

Taking into account the fact that QM/MM modeling will be performed and some semi-empirical QM calculations will take part in the active centre of the enzyme, validations of the semi-empirical methods were made. A computational model for peroxy form of FAD, one of the ligands in the active centre, was performed. The computational model used, represented 5-methyl-4 α -peroxy-10-H-isoalloxazine (Figure 5-3). The scan was performed around the C-O-O-H torsion (atoms 7, 30, 31 and 32). A good description of the O-O bond at the semi-empirical level was sought.

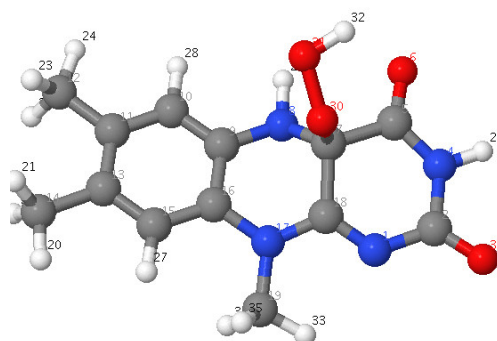


Figure 5-3. Structure of 5-methyl-4 α -peroxy-10-H-isoalloxazine

As well as performing a scan using B3LYP/6-31+G**, scans were also performed using AM1 and PM3 methods, and a comparison was then made of the three methods (Figure 5-4).

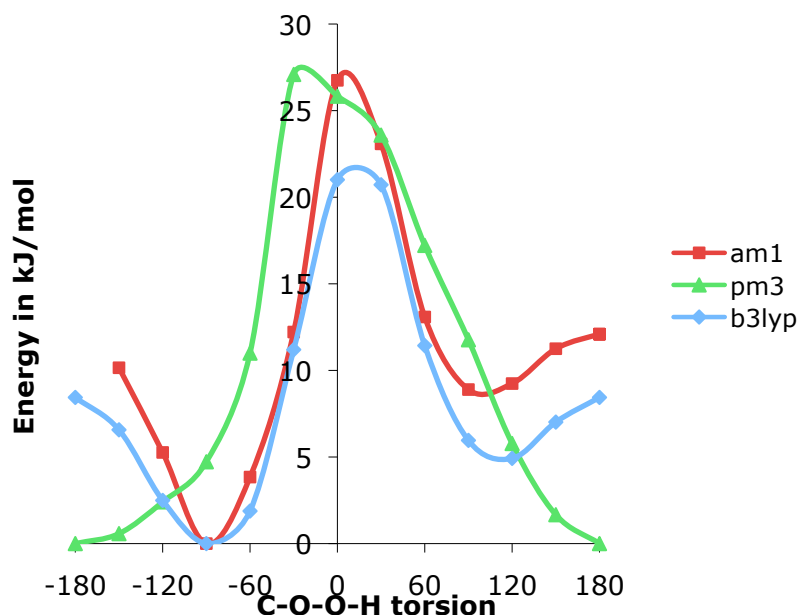


Figure 5-4. Scans with B3LYP, AM1 and PM3

The B3LYP scan results showed that there were two minima, at around 120° and -90° and maxima at around 0° and 180° of rotation around the O-O bond of the C-O-O-H torsion. Minima and maxima at the AM1 level were obtained at around 90° and -90°, and 0° and 180°, respectively. The PM3 scan revealed a minimum at 180° and a maximum at 30°. Of the two semi-empirical methods used, the AM1 method yielded results a better fit to the DFT scan results. It had two minima and two maxima similar to those of B3LYP. The PM3 scan revealed only one minimum and one maximum, that compared poorly with the B3LYP scan results. From the two maxima obtained at the AM1 and B3LYP levels, at 0° and 180°, the maximum at 0° had the higher barrier. It is probable that the structure passes from one minimum to the other on the PES (potential energy surface) through the 180° maximum. PM3 did not have a local maximum at 180°, but instead a minimum, hence it was excluded as a method to be used for QM/MM simulations

After a non-constrained optimisation, with starting structures of the two minima, a comparison between AM1 and B3LYP methods was made. The energy barriers and

the bond lengths confirmed that AM1 gives results similar to those obtained with DFT (Table 5-1).

	B3LYP	AM1	B3LYP	AM1
C-O-O-H torsion minimum after an optimisation	111.4°	103.5°	−84.6°	−82.9°
	1.44 Å	1.28 Å	1.44Å	1.28Å
O-O bond length at the minimum	(111.4°)	(103.5°)	(−84.6°)	(−82.9°)
Energy barrier at 0°	16 kJ/mol	19 kJ/mol	21 kJ/mol	28 kJ/mol
Energy barrier at 180°	4 kJ/mol	3 kJ/mol	9 kJ/mol	12 kJ/mol
Relative energies of the minimum from the scans	5 kJ/mol	9 kJ/mol	0 kJ/mol	0 kJ/mol

Table 5-1. Energy barriers and bond lengths for AM1 and B3LYP

5.2.2 Preparation of the system for the QM/MM simulations

A sphere of 30 Å of water molecules was cut around the active site of chain A and the remaining water molecules were deleted (Table 5-2).

Chemical group	Number of atoms
Amino acids	16392
FADHOOH 521	88
FAD 1042	84
ZTRP 520 + ZTRP 1041	54
Cl [−] 519 + Cl [−] 1040	2
Na ⁺ 1043/ Na ⁺ 1044	2
Water molecules	11760
Total number of atoms	28382

Table 5-2. Setup of the system for QM/MM simulations

The atoms within a 20 Å radius of the flavin ring, the chloride and -CH₂NH₂/-CH₂NH₃⁺ side chain of Lys78 in chain A, were allowed to move, whilst the other atoms remained frozen during all the simulations (Table 5-3, Figure 5-5).

Residue	Number of atoms
Amino acids' active atoms	6689
FADHOOH 521	88
ZTRP 520	27
Cl-519	1
Na ⁺ 1043	1
Number of active waters	2757
Total number of active atoms	9563

Table 5-3. Active atoms in the protein

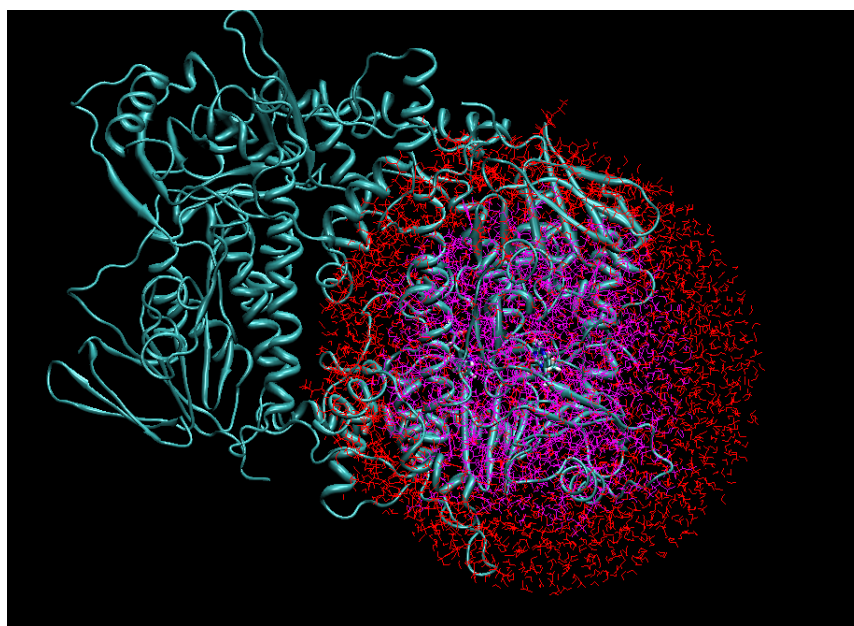


Figure 5-5. PrnA dimer setup system

A peroxy group was added to the FAD in the chain A, while the other FAD ligand was kept intact.

5.3 Results and Discussion

5.3.1 Nucleophilic attack of the chloride on FADHOOH

The first step of the mechanism included the formation of hypochlorous acid (HOCl) from the chloride anion and FADOOH. The reactants were initially optimised with AM1/AMBER03 and then refined at the TPSS/def2-TZVP+ level. The peroxyflavin ring, the chloride and three of the surrounding water molecules that were hydrogen-bonded to the chloride were selected for the QM region. This included 43 QM atoms without the link atom.

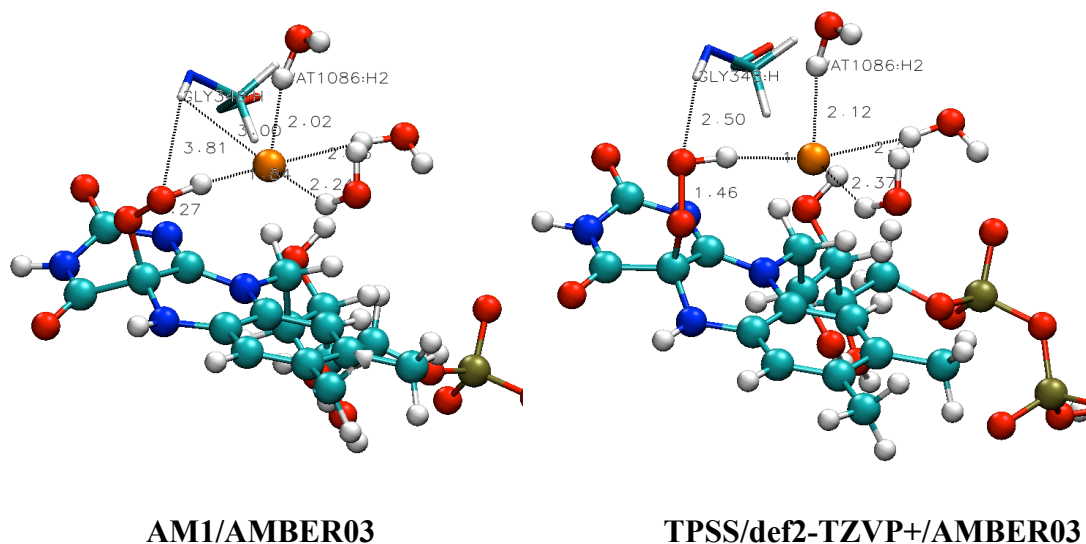


Figure 5-6. Optimised QM/MM FADHOOH and chloride reactants

The length of the hydrogen bonds obtained by the two different methods is given in the following table:

Hydrogen bonds	AM1/AMBER03	TPSS/def2-TZVP+/AMBER03
Cl ⁻ (519)···Wat1055	2.13	2.20
Cl ⁻ (519)···Wat1086	2.02	2.12
Cl ⁻ (519)···Wat1091	2.24	2.37

Table 5-4. Hydrogen bond interactions for the QM/MM optimised reactants in Å

For a reaction coordinate of the attack of the chloride on FADHOOH was chosen the difference of the Cl-O1 (FADHOOH) and O1-O2 (FADHOOH) distances was chosen.

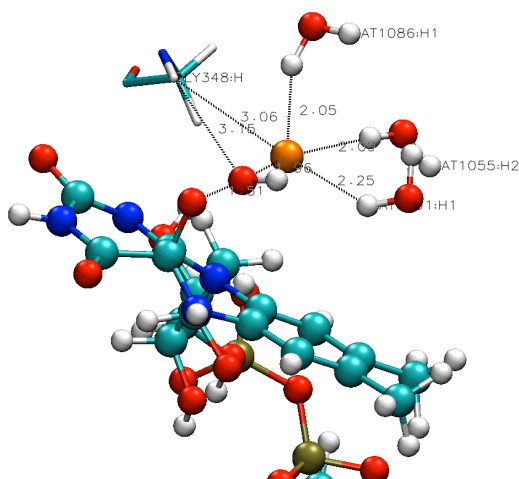


Figure 5-7. Optimised transition state structure for the attack of Cl519 on FADHOOH at the AM1/AMBER03 level

The energy barrier for the reaction, obtained using AM1/AMBER03 was 109 kJ/mol. The Cl-O1 (FADHOOH) and O1-O2 (FADHOOH) distances were 1.96 Å and 1.51 Å, respectively.

In the transition state the hydrogen bonds between the chloride and the surrounding waters remained almost unchanged compared to the reactants.

Hydrogen bonds	AM1/AMBER03
Cl ⁻ (519)···Wat1055	2.09
Cl ⁻ (519)···Wat1086	2.05
Cl ⁻ (519)···Wat1091	2.25

Table 5-5. Hydrogen bonds between the chloride and ambient water in the active site in the AM1/AMBER03 transition state structure

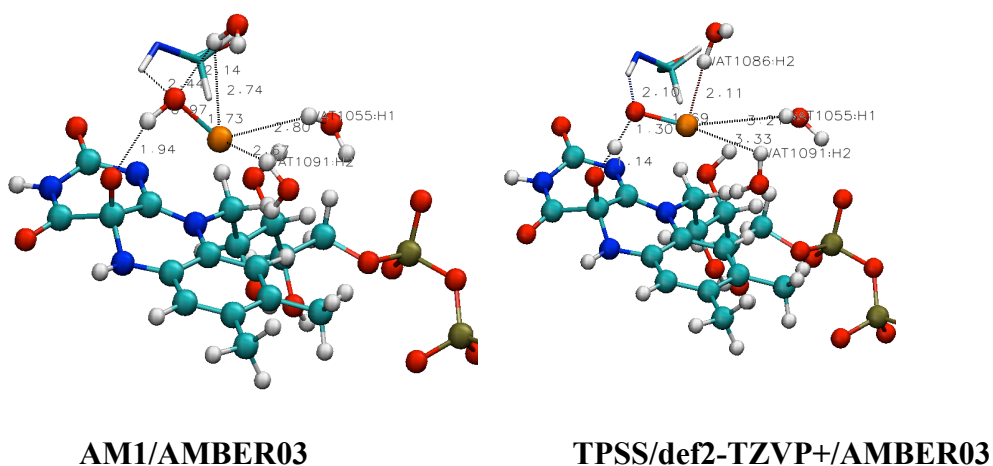


Figure 5-8. QM/MM optimised structures of the products of HOCl formation

The reaction energy calculated with AM1/AMBER03 was $\Delta E_r = -48$ kJ/mol. After refining the structure with TPSS/def2-TZVP+/AMBER03, the reaction energy altered to $\Delta E_r = 21$ kJ/mol.

The strong interaction between the hypochlorous acid and the FADO^- led to a proton transfer in the DFT/AMBER03 product. The hypochlorous acid reacted with the FADO^- base and a proton was transferred to the stronger FADO^- base. The distance between O2 (FADO^-) and the chloride was close to the normal O-H bond length- 1.14 Å. Between the hypochlorite anion and FADOH, a 1.3 Å hydrogen bond was formed. In the case of DFT/AMBER03, two processes occurred- the formation of HOCl and the base-acid proton transfer between FADO^- and HOCl.

The hydrogen bonds in the TPSS/def2-TZVP+/AMBER03 optimised structure, compared to the AM1/AMBER03 structure, are presented in the following Table 5-6:

Hydrogen bonds	AM1/AMBER03	TPSS/def2-TZVP+/AMBER03
Cl ⁻ (519)···Wat1055	2.80	3.21
Cl ⁻ (519)···Wat1086	2.74	2.11
Cl ⁻ (519)···Wat1091	2.87	3.07

Table 5-6. Hydrogen bond distances in the QM/MM simulated products

The change in the hydrogen bond lengths from the reactants, through the transition state (AM1/AMBER03) to the products was due to a change in the electrostatic interactions of the negatively charged chloride reactant to the neutral HOCl product.

5.3.2 Water networks and protonation state of Lys78

The protonation state of the Lys78 residue could not be determined unambiguously from the MD simulations. This was the reason for performing the simulations with both the LYN and LYP systems. It was found that in the protein three water molecules (Wat1076, Wat1103 and Wat3963) were connected to both residues, Lys78 and Glu345 through hydrogen bonds. There were two water chains, one that contained one water molecule (Wat3963), and the other containing two water molecules (Wat1076 and Wat3963) (Figure 5-9).

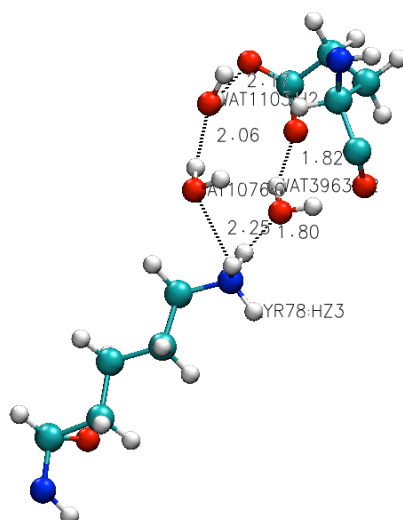


Figure 5-9. Water networks connecting Lyp78 and Glu345, after an optimisation at the AM1/AMBER03 level

Proton transfer can occur through these two water chains and result in a change the protonation state of the residues connected by the hydrogen bonded waters. The proton transfer through the water networks was studied, optimising the structures with one of the two residues of the protein, one protonated and the other non-protonated. The net reaction between the residues could be represented as follows:

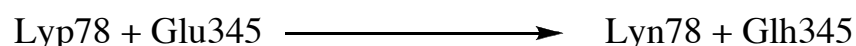


Figure 5-10. Net reaction of proton transfer between Lyp78 and Glu345

The reaction energies were first calculated initially using AM1/AMBER03, and then refined at the TPSS/def2-TZVP+/AMBER03 level. The QM region consisted of $\text{CH}_2\text{NH}_3^+/\text{CH}_2\text{NH}_2$ and $\text{CH}_2\text{COO}^-/\text{CH}_2\text{COOH}$ side chains of lysine and glutamate/glutamic acid and three water molecules being, 22 atoms in total. The reaction energies obtained at the DFT level gave preference to the system with protonated lysine and glutamate (Table 5-7). The reaction energies for the system with one water molecule suggested that the direction of the reaction is towards the non-protonated lysine. The DFT reaction energy was endothermic, but still low enough to make it plausible that the proton transfer can occur. The reaction energies

obtained for the system with two waters in the network suggested that proton transfer through that chain would not occur. The protonation state of lysine 78 is a question of interest because the non-protonated lysine can form a chloramine with HOCl, which can be the actual chlorinating agent. The protonated lysine can form a hydrogen bond with this residue.

	Network with Wat3963 in kJ/mol	Network with Wat1076 and Wat1103 in kJ/mol
AM1/AMBER03	-38	24
TPSS/AMBER03	50	275

Table 5-7. Reaction energies for the protonation of Lys78 with AM1/AMBER03 and TPSS/def2-TZVP+/AMBER03

The difference in energies of the chains consisting of two waters is due mainly to the change in the bond energy component. At the AM1/AMBER03 level the bond energy is 14 kJ/mol and it changes to 105 kJ/mol when the method was switched to DFT/AMBER03. In the network with one water, the bond energy is -4 kJ/mol with AM1/AMBER03, and -0.3 kJ/mol with DFT/AMBER03.

5.3.3 Attack of the hypochlorous acid on the Lys78 residue

The system setup for the attack of the hypochlorous acid on the lysine residue started with positioning the hypochlorous acid at different places around the lysine residue in the LYN system. The FADO⁻ formed in the previous step was saturated with a proton to FADHOH. The setup started with eight different positions. The structures were cleaned up with optimisation of the hypochlorous acid in a frozen environment. The eight trial structures were optimised within a free environment and collapsed to 3 different structures. Pulling the chloride of the hypochlorous acid towards the lysine residue to form the chloroammonium form of lysine (lysine78-NH₂Cl⁺) led to two adduct structures, complexes between the acid and the protein residue (Figure 5-11).

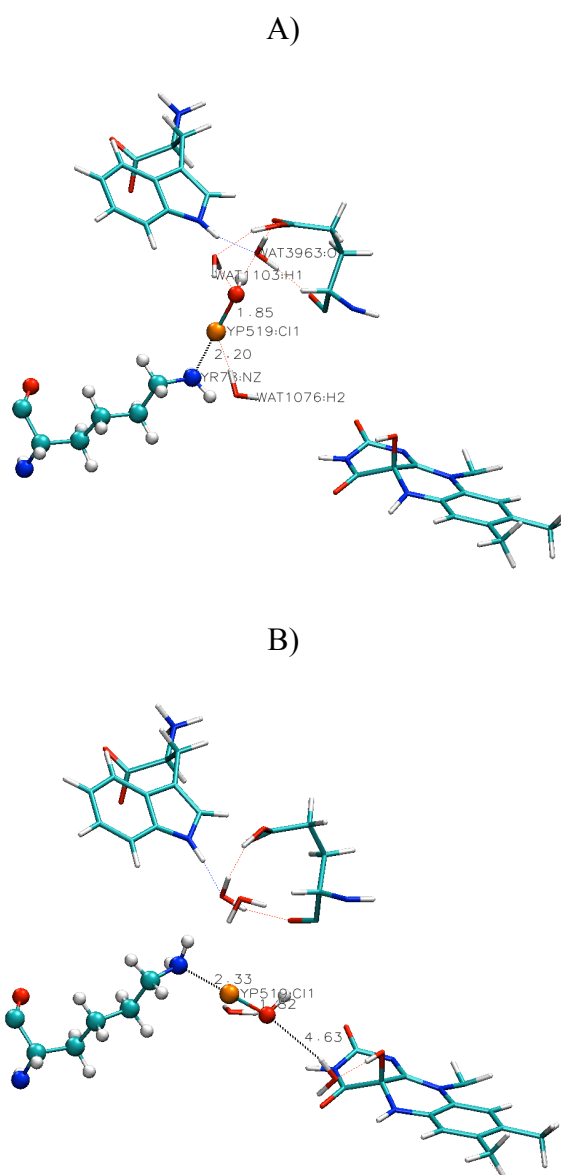


Figure 5-11. Adduct structures of the hypochlorous acid attacking Lyn78 with TPSS/def2-TZVP+/AMBER03

In these Lyn78- hypochlorous acid complexes the hypochlorous acid is situated on two opposite sides of the Lyn78. One of them is in the tunnel connecting the FADHOOH co-factor and the substrate (**A**), and the other is hydrogen-bonded to the water molecules connected to Glh345 (**B**) (Figure 5-11). The bond distances Cl-N (Lys78) and Cl-O (HOCl519) were 2.20 Å and 1.85 Å respectively, in complex **A**, and 2.33 Å and 1.82 Å respectively, in complex **B**. The resulting complexes are thermodynamically less stable than the reactants of free HOCl and LYN78 before

pulling of the chloride started by 65 kJ/mol for complex **A** and 21 kJ/mol for complex **B** at the TPSS/def2-TZVP+/AMBER03 level.

5.3.4 Formation of the Lyn78-chloramine intermediate

The products of Lyn78-chloramine were formed after a proton transfer to the hydroxide-leaving group of HOCl. In both setups, the source of the proton was different. The hydroxide group in structure **A** received a proton from Glh345. The proton that was transferred during the LYN78/LYP78 protonation state comparison was brought back from Glh345. DFT (TPSS) and a small basis set (def2-SVP+) were used to describe the process. The QM region included the -CH₂NH₂ and -CH₂COOH side chains of Lyn78 and Glh345, the hypochlorous acid, Wat1076, Wat1103 and Wat3963. For the reaction coordinate the difference of Cl-O (HOCl519) and O (HOCl519)-H (Wat1103) distances was chosen. The reaction energy for the proton transfer was 8 kJ/mol. After the hydroxide was protonated, the products of the reaction were formed- (chloroammonio)lysine and water. The structure was reoptimised at the TPSS/def2-TZVP+ level (Figure 5-12).

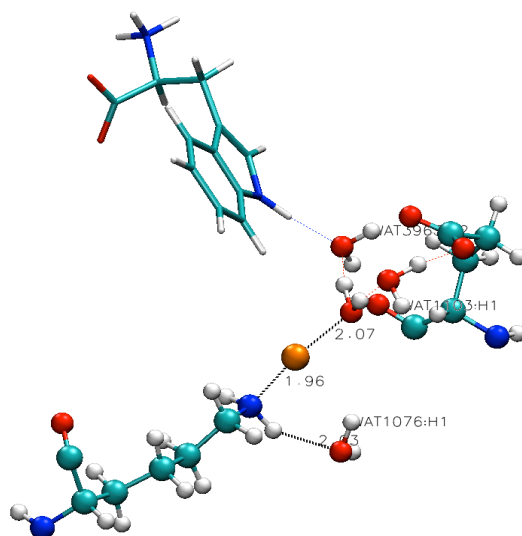


Figure 5-12. Product **A** of the HOCl attack on Lyn78 with TPSS/def2-TZVP+/AMBER03

The Cl-N (Lyn78) distance of the product was 1.96 Å and the Cl-O (HOCl519) distance- 2.07 Å. The estimated barrier for the reaction was 16 kJ/mol.

Continuing to take steps along the same reaction coordinate after the formation of the (chloroammonio)lysine led to formation of chloroamine. The (chloroammonio)lysine was deprotonated and the proton was transferred through Wat1076, Wat4965 from the hydroxide leaving group of HOCl and Wat1103, to the Glu345. The structure was again optimised with the def2-TZVP+ basis set (Figure 5-13).

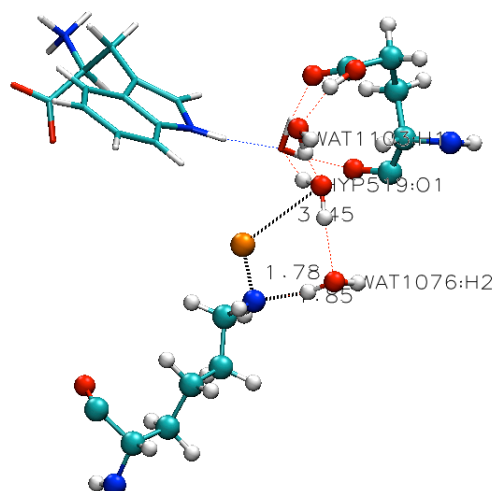


Figure 5-13. Optimised QM/MM structure in PrnA of the (chloroamino)lysine product **A** with TPSS/def2-TZVP+/AMBER03

The relative energy of the (chloroamino)lysine to (chloroammonio)lysine at TPSS/def2-SVP+/AMBER03 was -70 kJ/mol. The C-N (Lyn78) distance was 1.78 Å and the bond between the chloride and the oxygen in the hypochlorous acid was fully broken to 3.45 Å. (Chloroamino)lysine is more stable than (chloroammonio)lysine but it is not a viable electrophile. The products of the attack on tryptophan by (chloroamino)lysine would be 7-chlorotryptophan and (chloroamino)lysine anion (lysine-NH⁻) and the latter one is very unstable. That was the reason why the (chloroamino)lysine was ruled out as being the chlorinating agent.

For a source of protonation of the hydroxide leaving group in complex **B**, built from a water molecule hydronium cation was chosen. The FAD co-factor/Cl⁻ active site is connected to the bulk water outside of the protein and there is a channel through which the chloride can access the FAD-binding pocket. On this basis, it can be ascertained that it is possible for a proton to be delivered to the binding site in order to protonate and form a hydronium ion from a water situated near to the complex B. The proton can be transferred to the hydroxide leaving group and the products, Lyn-NH₂ +

2H₂O, can be formed. The structure of complex B and the hydronium was optimised at the AM1/AMBER03 level. After this optimisation the structure remained almost unchanged (Figure 5-14).

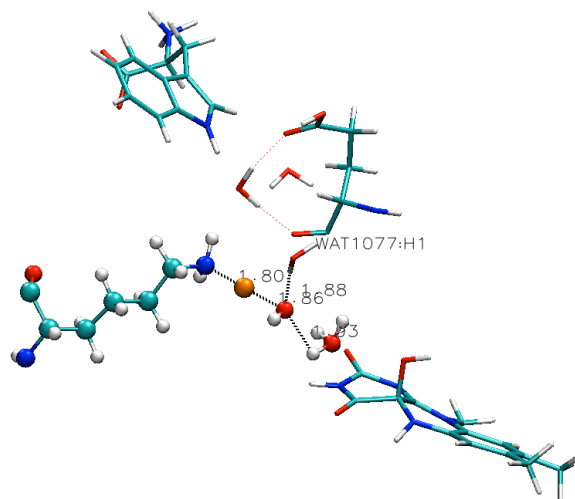


Figure 5-14. Optimised product **B** of (chloroammonio)lysine in PrnA with AM1/AMBER03

The further optimisation at the TPSS/def2-TZVP+/AMBER03 level led to the simultaneous proton transfer from the hydronium cation to the hypochlorous acid's hydroxide and full formation of the chloroammonium form of Lyn78 and water (Figure 5-15).

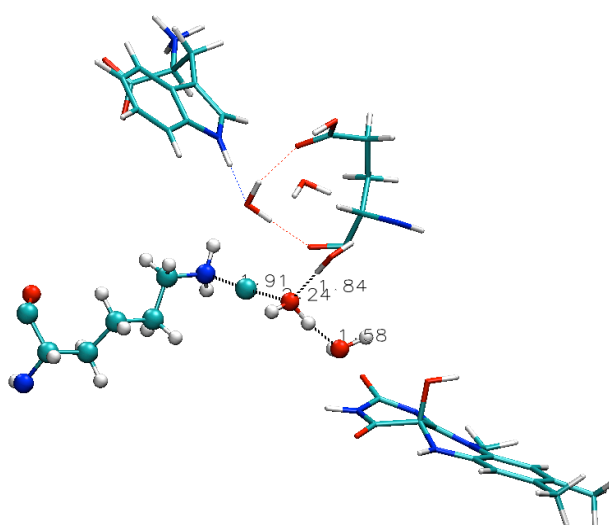


Figure 5-15. Optimised product **B** of (chloroammonio)lysine in the enzyme with TPSS/def2-TZVP+/AMBER03

5.3.5 Formation of the Tryptophan σ - complex

For the formation of the tryptophan σ -complex in the enzyme, two separate mechanisms were considered. In one of them the chlorinating agent was the (chloroammonio) lysine, and in the other one the electrophile was the hypochlorous acid, hydrogen-bonded to the Lyp78 residue (Reactions 4 and 5 of Figure 5-2)

5.3.5.1 (Chloroammonio)lysine78 as chlorinating agent

The optimisation of structure **A** with the (chloroammonio)lysine, was carried out by using TPSS/def2-TZVP+ with a different QM regions. It included the (chloroammonio)lysine and the substrate tryptophan, led to a change in the bond lengths of the chloroammonio group. The Cl-N (Lyn78) bond length changed from 1.96 Å to 1.70 Å (Figure 5-16).

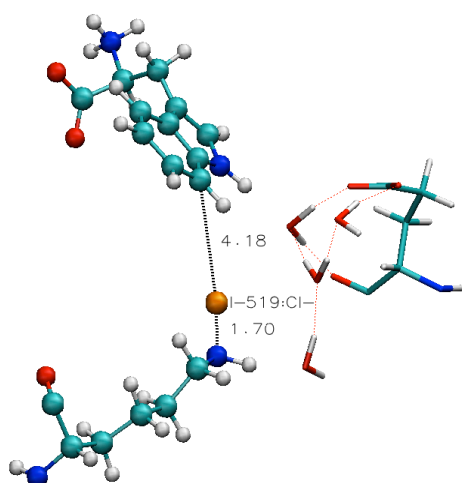


Figure 5-16. Starting QM/MM structure of the electrophilic attack on tryptophan at the DFT(TPSS/def2-TZVP+)/AMBER03 level

The rotation of the N-Cl (Lys78) of the chloroammonio form of Lyn78 in the QM/MM structure **B** gave the same starting structure for the attack on the substrate as system **A**. The rotational energy $\Delta E_{\text{TPSS/def2-TZVP+}} = -14$ kJ/mol suggested that the conformer with chloride positioned towards the substrate side of the lysine residue was the preferred one.

For the electrophilic attack a reaction coordinate of the difference of Cl-N (Lys78) and Cl-C7 (ZTRP520) distances was chosen. The calculations were performed with AM1/AMBER03. In the first steps conformational changes of ZTRP and (chloroammonio)lysine were observed. The two moieties approached each other. The arm of the lysine residue and the indole ring moved in directions that shortened the distance between them: $d(\text{Cl-N (Lys78)}) = 1.89 \text{ \AA}$ and $d(\text{Cl-C7(ZTRP)}) = 1.86 \text{ \AA}$.

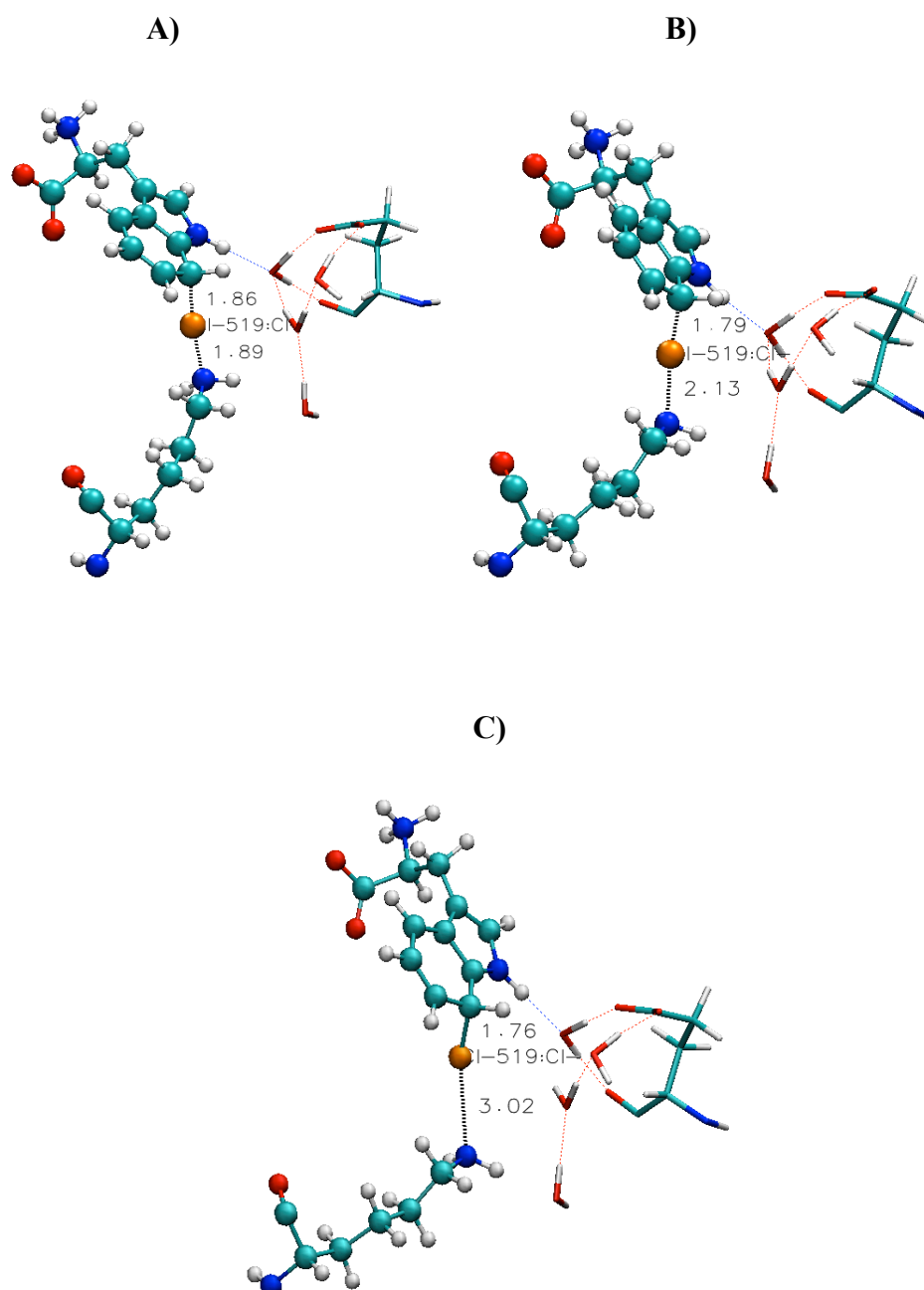


Figure 5-17. AM1/AMBER03 optimised structures of A) the (chloroammonio)lysine-tryptophan complex; B) the transition state of the (chloroammonio)lysine + tryptophan reaction and C) the products (the σ -complex of 7-chloroindole and Lys78)

The reaction energy was $\Delta E_r = -38$ kJ/mol, with a barrier of $\Delta E^\ddagger = 7$ kJ/mol. The low barrier could be explained by the fact that (chloroammonio)lysine is a positively charged residue and is a good electrophile.

5.3.5.2 Hydrogen-bonded hypochlorous acid as chlorinating agent

The second possible mechanism of chlorination was chlorination by the hypochlorous acid, hydrogen-bonded to the Lyp78. The hydrogen bond was proposed to orientate the acid in a preferable position for the attack on C7 of the indole ring of tryptophan. In the reactants the oxygen of hypochlorous acid was at a distance of 2.43 Å from the HZ3 hydrogen of Lyp78 and $d(\text{Cl}-\text{C7}(\text{ZTRP})) = 2.78$ Å after the optimisation with AM1/AMBER03. These distances changed to 2.62 Å and 2.70 Å at the DFT/AMBER03 level (Figure 5-18). The QM region included the side chain of the lysine, the hypochlorous acid and the tryptophan, 36 atoms in total.

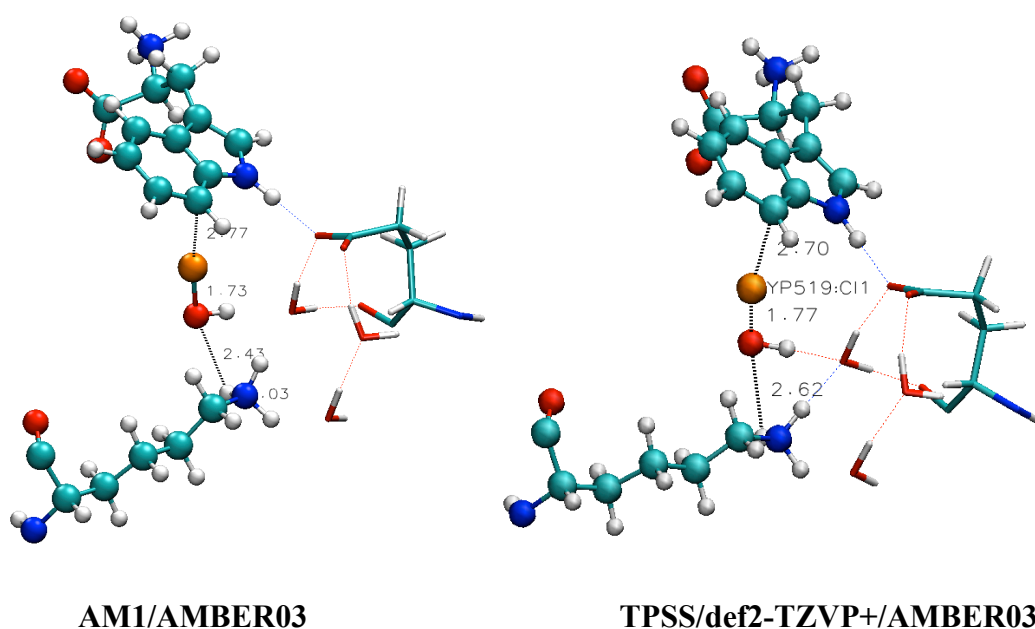


Figure 5-18. QM/MM Optimised Lyp78, HOCl and tryptophan and chloride reactants

The transition state and the products were found with a reaction coordinate of the difference of the distances N-HZ3 (Lyp78) and Cl-C7 (ZTRP). In the transition state with AM1/AMBER03, the proton of Lyp78 was situated in the middle between the Lyp78 residue and the oxygen of HOCl, at distances of 1.28 Å and 1.27 Å. The Cl-C7 (ZTRP) distance was 1.98 Å (Figure 5-19). The activation energy for the HOCl attack on tryptophan, with AM1/AMBER03, was $\Delta E^\ddagger = 97$ kJ/mol.

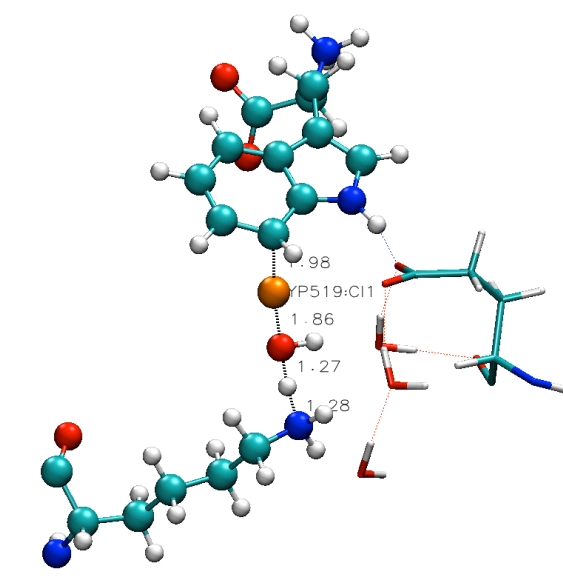


Figure 5-19. Optimised transition state structure of the attack of HOCl on tryptophan at the AM1/AMBER03 level

In the products, Cl-C7 bond and tryptophan σ - complex were formed, the hydroxide leaving group from HOCl was protonated by Lyp78, and these compounds were transformed to water and Lys78. In the products, a significant difference between the two QM methods in the QM/MM calculations was found in the Cl-C7 (ZTRP) bond length obtained. In the DFT/AMBER03 product it was 0.1 Å longer (1.85 Å) than in the AM1/AMBER03 product, where it was 1.75 Å (Figure 5-20). The reaction energy obtained with AM1/AMBER03 was $\Delta E_r = 9$ kJ/mol. The reaction energy calculated at the DFT/AMBER03 level was $\Delta E_r = 49$ kJ/mol.

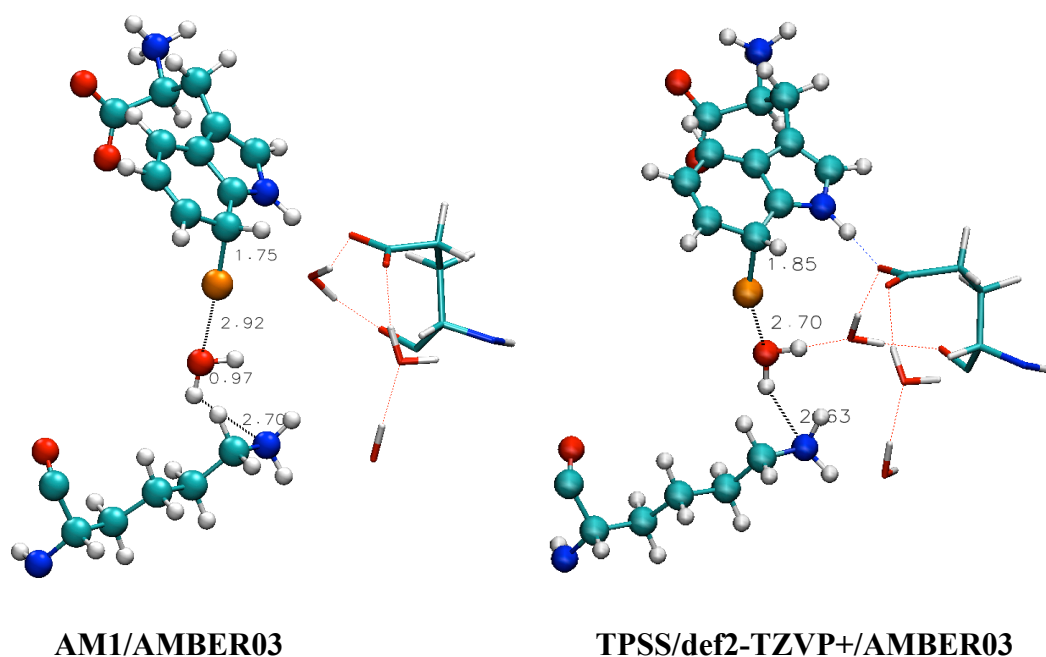


Figure 5-20. AM1/AMBER03 optimised structures of the products of the attack of HOCl on tryptophan

5.4 Conclusions

The last snapshot of the MD simulations of the LYP dimer with bound FAD was used as starting structure for the QM/MM calculations. The peroxy group of FADHOOH was built by hand. The system setup consisted of cutting a 30 Å sphere of water molecules around the active site of chain A where the chloride had a small deviation from the X-ray structure (0.90 Å) after the MD simulations. The active region, where the atoms were allowed to move, was in a diameter of 20 Å around the active centre of chain A and all the other atoms were kept frozen. The total number of atoms in the system was ~28 000 with ~10 000 of them active.

The methods, which were used for the QM/MM modelling were AM1/AMBER03 first and then the structures, were reoptimised with TPSS/def2-TZVP+/AMBER03. The semi-empirical method was validated after comparison of C-O-O-H torsion scan results in peroxyflavin to a B3LYP scan. The investigated reactions were: 1) formation of the hypochlorous acid through the nucleophilic attack of the chloride anion on FADHOOH; 2) proton transfer between the Lyp78 and Glu345 residues

forming the products Lyn78 and Glh345; 3) formation of the chloramine (LynNH₂Cl⁺) from Lyn78 and HOCl; 4) chlorination of the substrate ZTRP considering two separate mechanisms: the direct attack on ZTRP of hypochlorous acid hydrogen bound to Lyp78 and the attack on the substrate by the electrophile LynNH₂Cl⁺.

The reaction energy ΔE_r associated with the formation of hypochlorous acid from FADHOOH and chloride in the enzyme was relatively small with the combined QM/MM methods, -48 kJ/mol at the AM1/AMBER03 level with products FADHO⁻ and HOCl and 21 kJ/mol at the DFT/AMBER03 level with final products FADHOH and ClO⁻. The calculated barrier ΔE^\ddagger of the reaction with AM1/AMBER03 was 109 kJ/mol.

The residues in the protein playing an important role for the mechanism of the halogenation Lyp78 and Glu345 are connected via hydrogen bound water molecules. Proton transfer between the two residues is possible through water chains consisting of one and two water molecules. The reaction energies ΔE_r for the process of transferring a proton from Lyp78 (protonated lysine) to Glu345 (glutamate) at the DFT/AMBER03 level were 24 kJ/mol for the transfer of a proton through one water molecule against 275 kJ/mol when the transfer included two water-molecule chain. The proton transfer via a chain with one water molecule was disfavoured but possible. The likely protonation state of the Lys78 residue is the protonated lysine but in order to conduct one of the mechanisms of chlorination with LynNH₂-Cl⁺, Lyn78 was needed as a reactant for an attack of HOCl on the lysine residue.

For the mechanism of chlorination with Lyn78-NH₂Cl⁺, the nucleophilic agent was modelled by placing HOCl near Lyn78. As a result two adducts **A** and **B** were formed. The formation of the **B** adduct of Lyn78-NH₂Cl⁺ and HOCl at DFT/AMBER03 level, situated in the tunnel connecting FADHOH and ZTRP, was less exothermic (21 kJ/mol) than the formation of the **A** adduct of the same reactants (65 kJ/mol) but with HOCl, positioned near the substrate. The inclusion of water chains and Glh345 in the QM region invoked the possibility of proton transfer from Glh345 to OH⁻ leaving group of HOCl and a formation of water molecule from it. The reaction energy ΔE_r of the product of adduct **A** [Lyn78NH₂...Cl...OH₂]⁺ at the

DFT/AMBER03 level was 8 kJ/mol with a barrier ΔE^\ddagger of 16 kJ/mol. The deprotonation of this intermediate was exothermic with $\Delta E_r = -70$ kJ/mol and rendered Lyn78NH-Cl as a product. The newly formed neutral (chloroamino)lysine is not a good electrophile and was not used as chlorinating agent of ZTRP. The protonation of the OH⁻ leaving-group of HOCl in adduct **B** was by manually added proton to nearby water, a hydronium was built, assuming that the active centre is connected to the bulk water outside of the protein from where a proton could be delivered to the inside. The formation of the product [Lyn78NH₂...Cl...OH₂]⁺ of adduct **B** occurred spontaneously. The rotation about the C-N bond in LynNH₂Cl⁺ turned the product **B** to product of **A** with $\Delta E_r = -14$ kJ/mol. Scheme **B** prevails over scheme **A** as possible pathway for the LynNH₂Cl⁺ formation. The reaction energy ΔE_r was 21 kJ/mol for the adduct **B** formation and the exothermic rotation (-14 kJ/mol) of the C-N bond, which transformed it to **A**, gave an overall energy of the process of 7 kJ/mol. Another advantage of this scheme is that HOCl was positioned in the tunnel connecting the Lyn78 and FADHOH, which is the first position the acid would occupy leaving FADHOH.

The reaction energies ΔE_r for the formation of the σ - complex with two different chlorinating agents showed that the better chlorinating agent would be the (chloroammonio)lysine with $\Delta E_r = -38$ kJ/mol compared to the hypochlorous acid, hydrogen-bonded to Lyp78 with $\Delta E_r = 9$ kJ/mol, all calculated with AM1/AMBER03. The energy barriers ΔE^\ddagger for both mechanisms showed again a preference for the mechanism with (chloroammonio)lysine attacking the substrate, with a barrier of 7 kJ/mol, against the 97 kJ/mol barrier of the HOCl attack on tryptophan at the AM1/AMBER03 level.

Both halogenating agents, hypochlorous acid (HOCl) and LynNH₂Cl⁺ (chloroammonio)lysine, have their pros and cons. The barrier for the chlorination with the HOCl hydrogen-bonded to Lyp78 was relatively high (97 kJ/mol). On the other hand, despite the low barrier of the chlorination with LynNH₂Cl⁺ (7 kJ/mol), it required endothermic deprotonation of Lyp78. A disadvantage of the LynNH₂Cl⁺ is its strongly exothermic deprotonation to the long-lived LynNHCl intermediate, which

may correspond to the experimentally detected long- lived intermediate. The energies of all the reactions for different mechanisms are listed below:

Process	AM1/AMBER03		DFT/AMBER03
	ΔE_r	ΔE^\ddagger	ΔE_r
Formation of HOCl	-48	109	21
Proton transfer from Lyp78 to Glu345 through Wat 3963	-38	-	24
Proton transfer from Lyp78 to Glu345 through Wat1076 and Wat1103	50	-	275
Formation of complex A from Lyn78 and HOCl	-	-	65
Formation of complex B from Lyn78 and HOCl	-	-	21
Formation of Lyn-NH ₂ Cl ⁺ after a proton transfer / Formation of adduct A [Lyn78NH ₂ ...Cl...OH ₂] ⁺	-	-	7
Formation of Lyn-NHCl from complex A after a proton transfer to Glu345	-	-	-70
Rotation of -NH ₂ Cl ⁺ group in Lyn78-NH ₂ Cl ⁺ / Reaction of turning of B into A	-	-	-14
Formation of the 7-chlorotryptophan σ - complex from Lyn-NH ₂ Cl ⁺ + tryptophan	-38	7	-
Formation of the 7-chlorotryptophan σ - complex from HOCl hydrogen bonded to Lyp78 + tryptophan	9	97	49

Table 5-8. Reaction energies and barriers relative to the free reactants in kJ/mol of all the steps along the mechanism of chlorination of tryptophan in PrnA using QM/MM methods

The results presented in this chapter could be used for experimental/theoretical comparison based on the TST (Transition State Theory). Naismith *et al.* reported

kinetic data for the chlorination of tryptophan in PrnA.⁶⁴ The rate constant would correspond mainly to the slowest reaction in the process of chlorination *i.e.* the formation of the hypochlorous acid from FADHOOH and Cl^- in the enzyme. Then the calculated free energy (ΔG^\ddagger) could be compared to the experimental data. Different functional and basis sets (QM methods), part of the hybrid QM/MM method could be tried and their accuracy can be evaluated by comparison to the experimental data.

Chapter 6 Pyridoxal derivatives: Structural and spectroscopic investigation

6.1 Introduction

Vitamin B₆ (pyridoxine) is a precursor of the two enzymatic co-factors- pyridoxal 5'-phosphate (PLP) and pyridoxamine-5'-phosphate (PMP) (Figure 6-1). Pyridoxine is the product of biosynthesis in bacteria, fungi and plants. Parasites and higher organisms obtain in vitamin B₆ from nutrients.¹²⁸

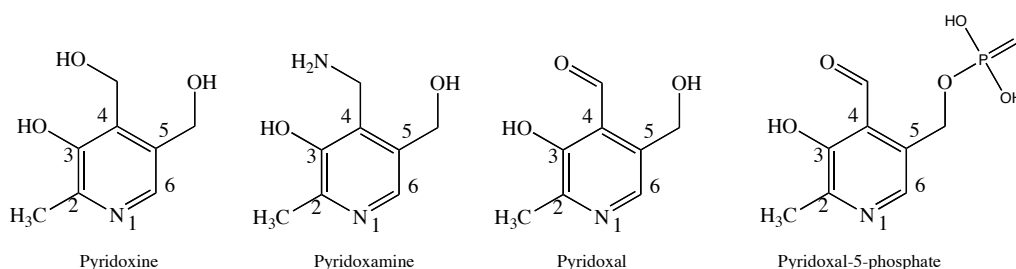
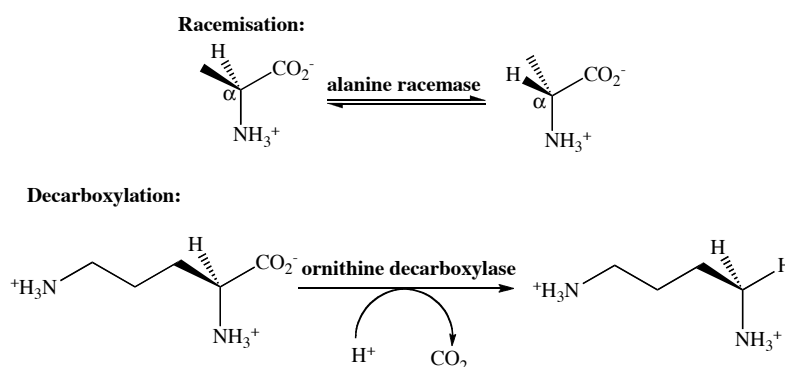


Figure 6-1. Vitamin B₆ (pyridoxine) and its derivatives

PLP participates in enzymatic reactions with amino acids, carbohydrates, lipids and nucleic acids as substrates. PLP plays the role of a co-factor in the enzyme-catalysed reactions of α -, β -, and γ -racemisation, decarboxylation, α -, β -, and γ -elimination and transamination (Figure 6-2). Inhibitors of PLP-enzyme interactions play a significant role in SHMT- cancer therapy and treatments for epilepsy and African sleeping sickness for example.¹²⁹



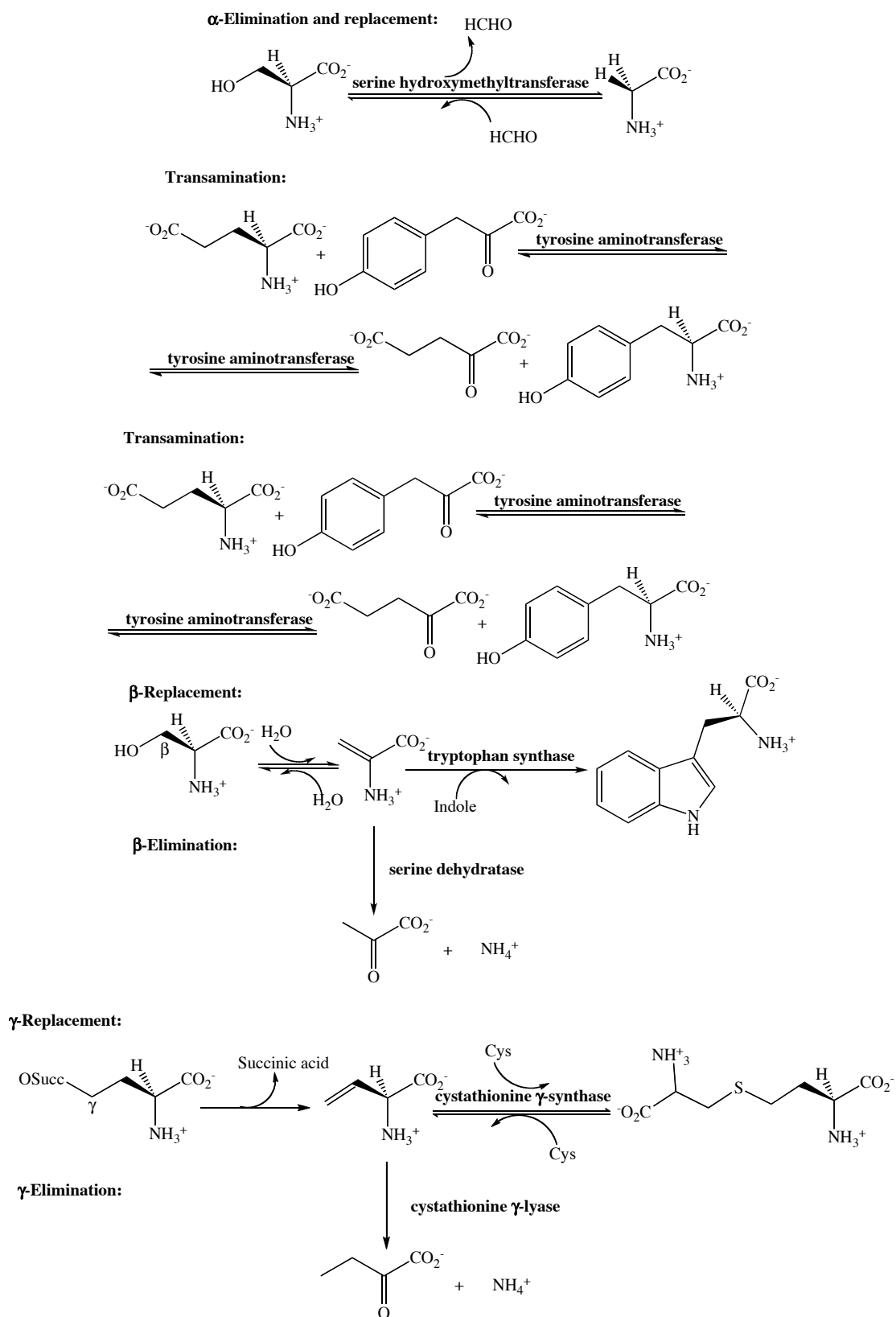


Figure 6-2. Reactions with PLP as a co-factor in variety of enzymes¹²⁹

A closer look at the specific mechanism of one of these reactions, shown above, is presented in Figure 6-3:

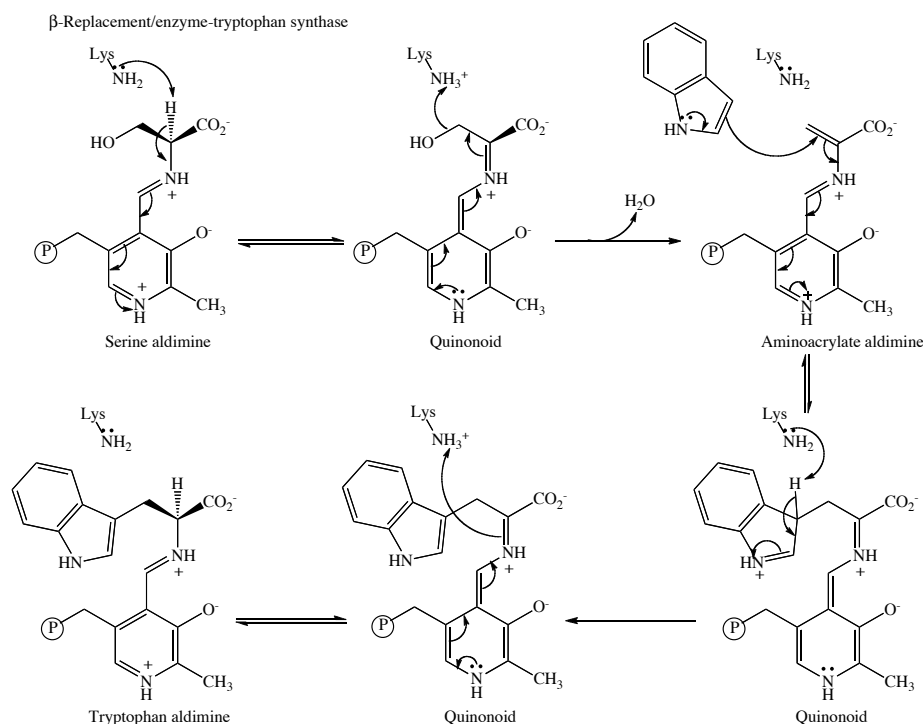


Figure 6-3. Mechanism of β-Replacement for tryptophan synthase¹²⁹

The general mechanism of the reactions involving PLP begins first with formation of a Schiff base with a residue from the protein, usually lysine. This Schiff base is called an internal aldimine. The incoming amino acid substrate replaces the lysine in a transaldimination reaction and a new aldimine is formed- externally this time. The deprotonation of the external aldimine by the lysine residue gives a quinonoid as an intermediate. Quinonoid can exist as one of three possible resonance structures in this intermediate (Figure 6-4). The negative charge is delocalized in the conjugated pi-system of the quinonoid.¹²⁹

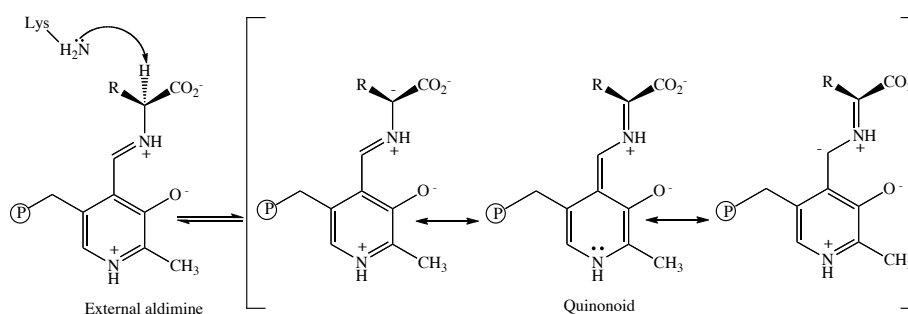


Figure 6-4. Resonance structures of the deprotonated external aldimine

After the formation of the quinonoid, the mechanism of the enzyme-catalysed reactions becomes more specific.

Metzler et al. measured the absorption spectra of the different protonated species of the non-enzymatic PLP and diethyl aminomalonate (Figure 6-5). The diethyl aminomalonate and PLP spectra gave a quinonoid peak at 460 nm, which is transformed into a band with a peak 330nm as the length of time increases.¹³⁰

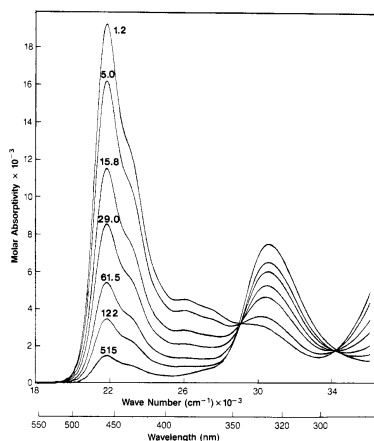


Figure 6-5. Spectrum of a mixture of 4.96×10^{-5} M PLP and 0.02 M diethyl aminomalonate at pH 6.2. The time in minutes is given by each curve¹³⁰

Alanine racemases are PLP-dependent enzymes that catalyse the racemisation of L-alanine to D-alanine. D-alanine is an amino acid, part of the peptidoglycan in bacteria. The peptidoglycan is a glycoprotein, a building block of the bacteria cell wall. The alanine racemases are the target of some newly discovered alanine-racemase inhibitors, which are used as antibacterial drugs. The inhibitor activity of β,β,β -trifluoroalanine in *S. typhimurium* and *B. stearothermophilus* has been investigated.¹ The mechanism was studied using $[1-C^{14}]$ -trifluoroalanine and a UV-Vis spectrum was measured during the experiments. The results of the spectrum showed a change in λ_{max} from 418nm (characteristic for the PLP Schiff bases) to 460-490 nm. The complete mechanism of inhibition of the alanine racemases is presented in Figure 6-6.

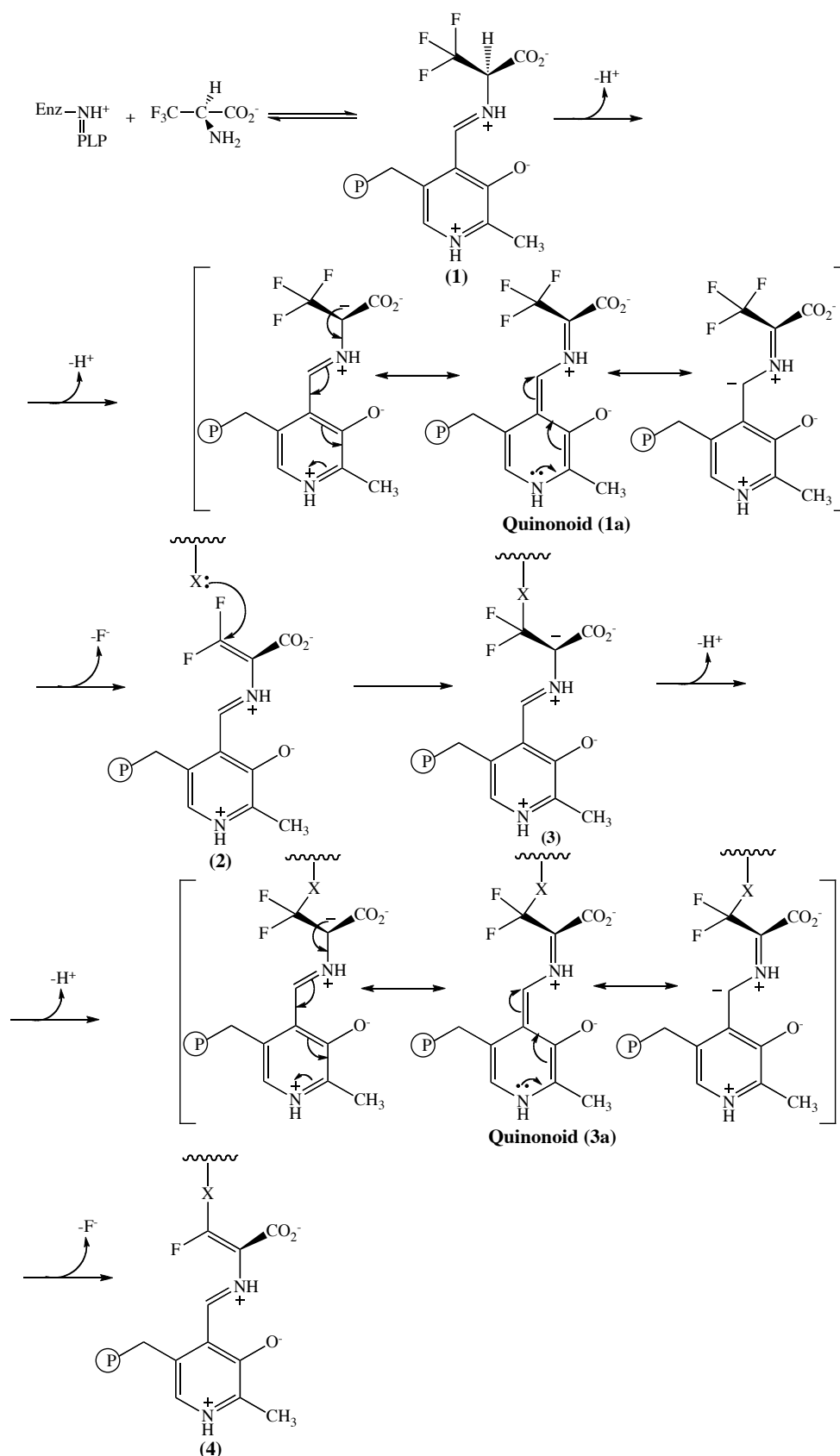


Figure 6-6. Mechanism of inhibition of alanine racemase involving β,β,β -trifluoroalanine¹

The addition of β,β,β -trifluoroalanine to the alanine racemase enzyme produces compound **(1)**- the external aldimine. The abstraction of the α -proton gives resonance structures **(1a)**, including the quinonoid form. The loss of F^- leads to a β -difluoro- α,β -unsaturated imine **(2)** and the consequent inactivation of the enzyme. This is followed by removal of another proton and formation of the quinonoid resonance forms **(3a)**. The elimination of the second fluoride anion is the final step in the formation of a stable enzyme-inhibitor complex **(4)** and inactivation of the enzyme.

The aims of the project were optimising and calculating the spectra of all the intermediates of the mechanism of inhibition of the alanine racemase; validation of the computed spectra after a comparison with the experimental data; calculating the deviation from the experiment; determining the wavelength in the absorption spectra of the each of the protonation states of the intermediates.

6.2 Computational methodology

The calculations were performed at the M06-2X/6-311+G(2df, p) level using the *Gaussian 09* package.¹³¹ M06-2X¹³² is a hybrid functional with double exchange (2X), and is parameterised only for non-metals. Zhao and Truhlar calculated the MUE (mean unsigned error) for the excitation energies of the 21 valence transition of various small molecules with different functionals. The M06-2X gave a satisfactory error of 0.34 eV.¹³² Optimisations and spectra calculations were performed in gas phase and solution using an IEFPCM continuum solvation model^{119,120} with $\epsilon_r=78.3553$ (corresponding to water solvent). In the TDDFT¹⁵ calculations, 50 excitations for singlet-excited states were requested.

6.3 Results and Discussion

The literature gives information about the spectra of pyridoxal 5'-phosphates in different protonation states.¹³³⁻¹³⁷ These spectroscopic data were used for validation of the method and for derivation of ΔE for further theoretical spectroscopic studies.^{138,139}

The quinonoid-Schiff bases show transitions in the visible part of the spectrum. The difference in the protonation state, tautomer forms or enzyme environment can cause changes in the spectral properties of the aldimine. Our aim was to study the influence of the structural tautomerisation and protonation states of all the compounds involved in the reaction of inactivation of the alanine racemase by β,β,β -trifluoroalanine.

6.3.1 Validation spectra

Depending on the pH, PLP can adopt different protonation states (Table 6-1). Several experimental groups have collected spectroscopic data.^{135,136} The UV spectra were measured in aqueous solutions at different pH. The initial optimisations of the structures and spectra were calculated in the gas phase in order to save their computational cost. The phosphate group in the di-anionic form was unstable and in order for it to be stabilised it was singly protonated.

The calculated and experimental spectra are presented in Table 6-1.^{135,136} The energy needed to promote an electron from the ground state to a given excited state is called “vertical” excitation energy, and it can be expressed in terms of the wavelength.

$$\Delta E = \frac{hc}{\lambda} \quad (6-1)$$

The same is valid not just for the absorption spectra but for the emission as well.

Based on the calculated and experimental excitation energies in the absorption spectra, extinction coefficients and the intensities it was possible to assign the calculated values to the experimental ones for the excitations. The assignment of a specific transition to a λ_{max} value was done by looking at the deviations and generating the orbitals. In the table each of the suggested assignments is highlighted in different colours. The λ_{max} values that are tabulated are those with oscillator strength (f) greater than 0.05. The cut-off λ_{max} was 165nm.

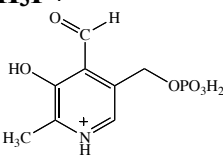
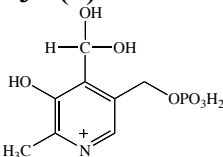
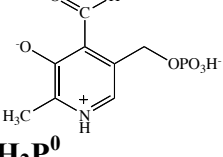
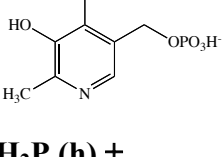
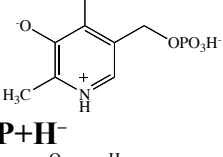
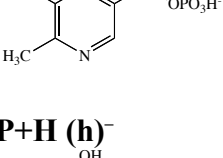
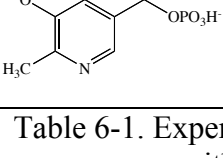
Compound (Protonation state)	pH	Experimental			Calculated		
		λ_{\max} (nm)	ϵ (l x mol ⁻¹ cm ⁻¹ x 10 ⁻³)	ΔE (Jx10 ⁻¹⁹)	λ_{\max} (nm)	f	ΔE (Jx10 ⁻¹⁹)
H₃P⁺ 	1.0	336 253	1.35 0.55	5.91 7.85	304 235 199 196 171 169	0.19 0.06 0.15 0.12 0.11 0.08	6.53 8.45 9.43 9.98 10.13 11.62
H₃P (h)⁺ 	1.0	294 220	6.61 3.39	6.76 9.03	261 195 186 175	0.21 0.16 0.16 0.19	7.61 10.19 10.68 11.35
H₂P[±] 	6.17 6.90	390 274 231	4.84 0.78 10.25	5.09 7.25 8.60	369 216 212 167	0.20 0.23 0.06 0.05	5.38 9.20 9.37 11.89
H₂P⁰ 	6.17 6.90	351 249 205	0.86 2.22 4.80	5.66 7.98 9.69	323 246 201 198 197 170	0.11 0.09 0.12 0.06 0.09 0.13	6.15 8.07 9.83 10.03 10.09 11.68
H₂P (h) ± 	3.15 6.17 6.90	322 251 213	1.32 0.84 6.27	6.17 7.91 9.33	305 204 203 168	0.17 0.21 0.12 0.08	6.52 9.74 9.79 11.82
P+H⁻ 	6.90 8.35 8.70 14	391 261 229	6.21 2.74 12.41	5.08 7.61 8.67	350 228 218 191 169 168	0.23 0.08 0.08 0.09 0.13 0.07	5.68 8.71 9.11 10.40 11.75 11.82
P+H (h)⁻ 	8.35 8.70 14	302 239	0.58 0.54	6.57 8.31	285 236 198 177 174 172	0.13 0.07 0.12 0.09 0.10 0.06	6.97 8.41 10.03 11.22 11.41 11.54

Table 6-1. Experimental (aqueous solution)^{135,136} and calculated (gas phase) spectra with M06-2X/6-311+G(2df, p) of pyridoxal 5'-phosphate

The average values of $\Delta\Delta E$ for the assigned peaks were 0.51×10^{-19} J for all the first peaks and 1.12×10^{-19} J for the total assignments. The conversion between J and eV is: $1 \text{ eV} = 1.60217733 \times 10^{-19}$ J and $1 \text{ J} = 0.624150636 \text{ eV}$. It can be considered that the first value was more reliable compared to the other assignments.

Looking at the spectrum and the orbitals of the fully protonated PLP species H_3P (Figure 6-7 and Figure 6-8) made it possible to assign the transition of the electrons in the electronic spectrum.

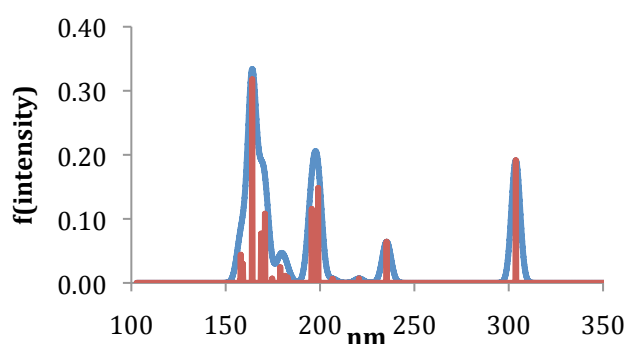


Figure 6-7. Calculated absorption spectrum of H_3P at the M06-2X/6-311+G(2df, p) level in gas phase

The population of the orbitals gives one an idea of the type of transitions that can be given with orbitals of the ground and different excited states. At Figure 6-8 are assigned the type of transitions to the first three peaks in the absorption spectrum of H_3P with oscillator strength greater than 0.5. The orbitals were generated using the M05-2X/6-311+G(2df, p) method in gas phase. The wavelength is given by λ , the oscillator strength is f , then the orbitals involved in the transition are given as numbers and notation, the weights of transition are the CI and the type of transition is given as σ , σ^* , π , π^* etc.

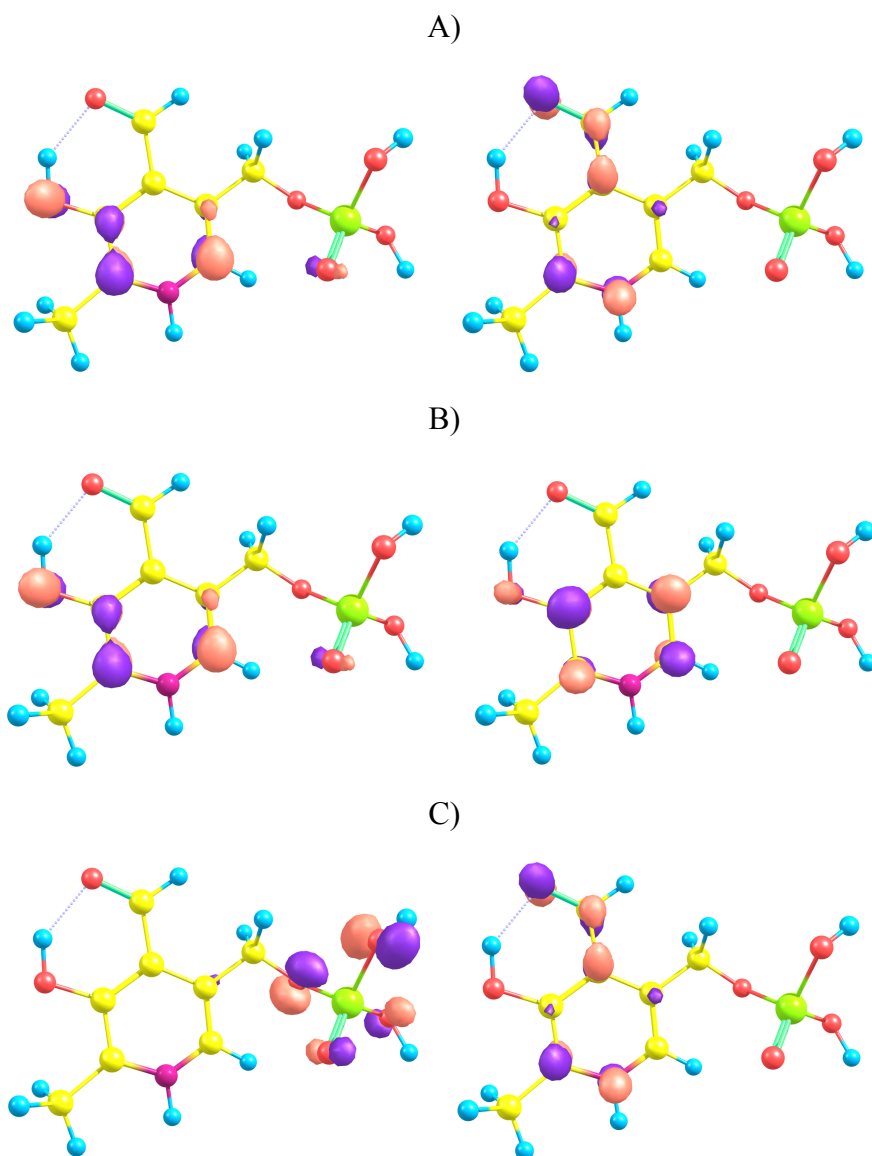


Figure 6-8. Molecular orbitals' representations of H_3P generated at the M05-2X/6-311+G(2df, p) level in gas phase: A) Excited state 2: $\lambda = 303.74$ nm, $f = 0.19$, $E = 4.08$ eV, $64 \rightarrow 65$ orbitals transition, $CI = 0.70$; B) Excited state 6: $\lambda = 199$ nm, $f = 0.15$, $E = 6.23$ eV, $64 \rightarrow 66$ orbitals transition, $CI = 0.50$; C) Excited state 7: $\lambda = 196$ nm, $f = 0.12$, $E = 6.34$ eV, $59 \rightarrow 65$ orbitals transition, $CI = 0.25$

The first two peaks in the spectrum of H_3P correspond to a HOMO-LUMO and a HOMO-LUMO+1 $\pi \rightarrow \pi^*$ transition. The third peak corresponds to a HOMO-5-LUMO $p \rightarrow \pi^*$. Charge transfer excitation is observed in this case. Charge transfer is the main issue when using TDDFT to calculate electronic spectra and in this molecule, substituting the phosphate group by a hydroxyl group can eliminate it. The phosphate group did not appear in the UV spectrum and it could be easily substituted without any impact on the absorption spectrum. Another issue was the stability of the

protonated species, which was extremely unfavourable in the gas phase, whereas the experimental data were measured using water as the solvent. This was the motivation to optimise the molecules in continuum water and in the gas phase and to calculate the spectra in the gas phase and in continuum water. The species with deprotonated hydroxyl group after the optimisation reached a minimum with a resonance structure with a carbonyl group. This fact was proved after measuring the bond lengths in the heteroaromatic pyridine ring. An example of this type of species is H_2P^\pm (Figure 6-9).

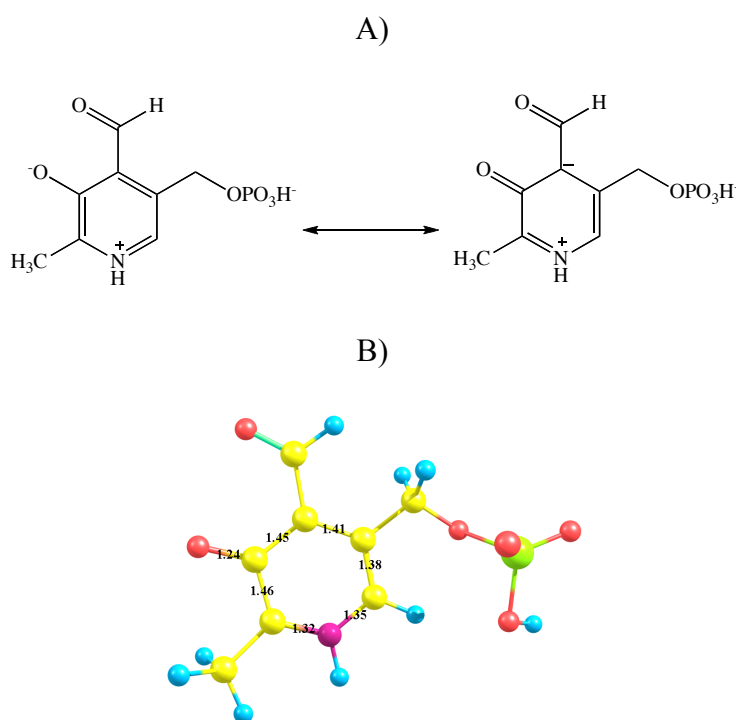


Figure 6-9. A) Resonance structures of H_2P^\pm ; B) A final structure after an optimisation of H_2P^\pm at the M06-2X/6-311+G(2df, p) level in gas phase

The experimental UV-spectrum that was available was for the aldehyde and hydrate forms of P^- . The two forms exist in equilibrium at a given pH and both appear in the UV spectrum (Figure 6-10).

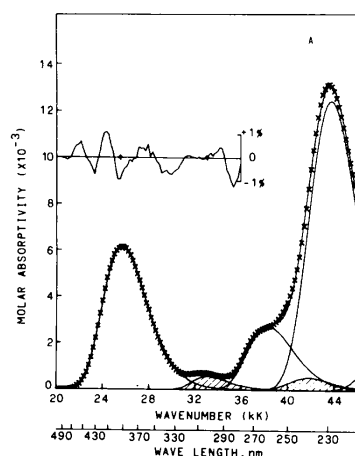


Figure 6-10. Experimentally measured spectrum of the most anionic form of PLP; the major band corresponds to the aldehyde and the small bands are the covalent hydrate¹³⁶

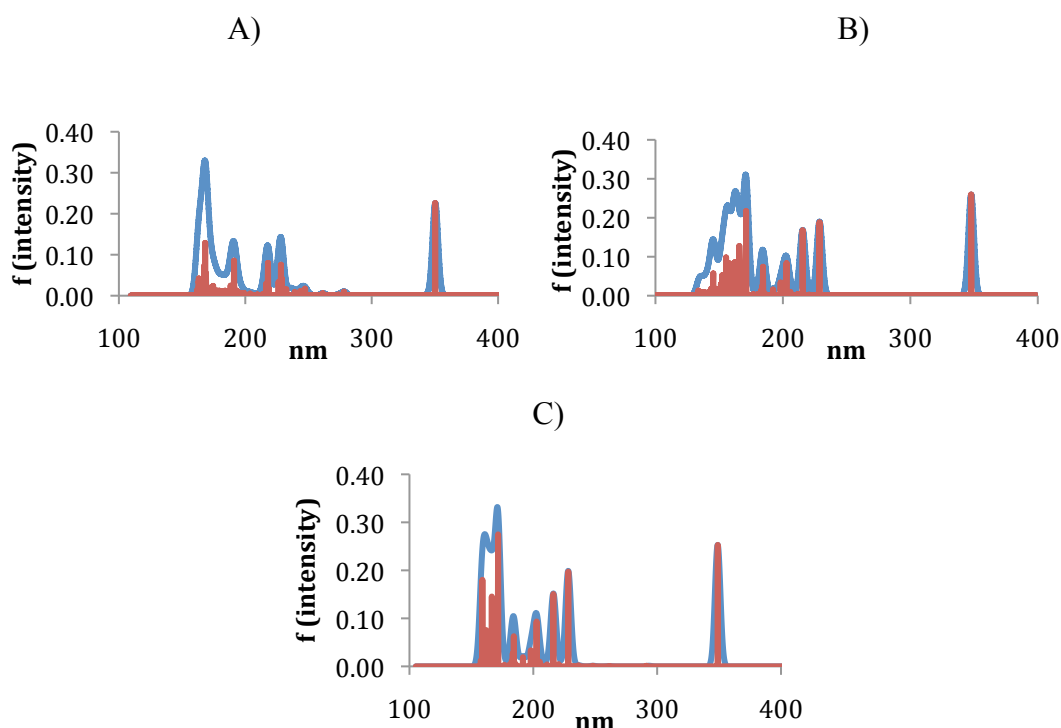


Figure 6-11. Calculated spectra of PLP or pyridoxal: A) Optimised PLP in gas phase with spectrum calculated in gas phase; B) Optimised pyridoxal in gas phase with spectrum calculated in IEFPCM water; C) Optimised pyridoxal in IEFPCM water and spectrum in IEFPCM water

The visual comparison between the experimental and calculated spectra of PLP and pyridoxal led to the conclusion it was necessary that the spectrum had to be calculated in water. The spectra with structures optimised in the gas phase and water did not differ drastically and to improve the computational efficiency the structures were optimised in the gas phase and their spectra were calculated in water.

Compound (protonation state)	Experimental spectra			Gas phase optimised structures with spectra calculated in IEFPCM water		
	λ_{\max} (nm)	ϵ (l x mol ⁻¹ cm ⁻¹ x 10 ⁻³)	ΔE (J x 10 ⁻¹⁹)	λ_{\max} (nm)	f	ΔE (J x 10 ⁻¹⁹)
H ₃ P ⁺	336	1.35	5.91	305	0.21	6.51
	253	0.55	7.85	241	0.07	8.24
				195	0.47	10.18
H ₃ P (h) ⁺				173	0.46	11.48
	294	6.61	6.76	256	0.23	7.56
	220	3.39	9.03	196	0.28	10.13
H ₂ P [±]				181	0.17	10.97
				174	0.09	11.41
	390	4.84	5.09	361	0.25	5.50
H ₂ P ⁰	274	0.78	7.25	218	0.45	9.11
	231	10.25	8.60			
	351	0.86	5.66	308	0.15	6.45
H ₂ P (h) [±]	249	2.22	7.98	236	0.16	8.41
	205	4.80	9.69	200	0.32	9.93
				174	0.16	11.41
P ⁻				167	0.38	11.89
	322	1.32	6.17	304	0.21	6.53
	251	0.84	7.91	233	0.11	8.52
P (h) ⁺	213	6.27	9.33	205	0.22	9.67
				172	0.07	11.54
				170	0.25	11.68
	391	6.21	5.08	348	0.26	5.71
	261	2.74	7.61	229	0.19	8.67
	229	12.41	8.67	216	0.17	9.19
				203	0.08	9.78
				184	0.07	10.79
				171	0.22	11.61
				166	0.13	11.96
				162		
	302	0.58	6.57	278	0.18	7.14
	239	0.54	8.31	233	0.16	8.52
				194	0.18	10.23
				190	0.2	10.45
				177	0.08	11.22
				176	0.21	11.28
				165	0.08	12.03

Table 6-2. Experimental^{135,136} and calculated spectra using with M06-2X/6-311+G(2df, p) method of pyridoxal in IEFPCM continuum water

The calculation of the $\Delta\Delta E$ values, derived as difference between the calculated and experimental ΔE , gave an average value for the first peaks of 0.63×10^{-19} J and 0.75×10^{-19} J for the total assignments. As in the gas phase the deviation between the

water pyridoxal assignments showed a smaller deviation from the experiment for the first peaks than the totalled values, which included an average of all the assignments. Compared to the gas phase PLP assignments of 0.51×10^{-19} J (first peaks) and 1.12×10^{-19} J (all the assignments) the water pyridoxal assignments gave not so good deviation in the first peak assignments but better in total. The overall assignments were more closely resembled the experiments. The experimental data that were used were measured in water, and calculating the spectra in water was logically the better approach

6.3.2 External aldimines' structures and spectra

The first step of the mechanism of inhibition of the alanine racemase by β,β,β -trifluoroalanine was the reaction between pyridoxal and β,β,β -trifluoroalanine.¹ The product of this step is the external aldimine Schiff base. The reaction can give six possible protonation states:

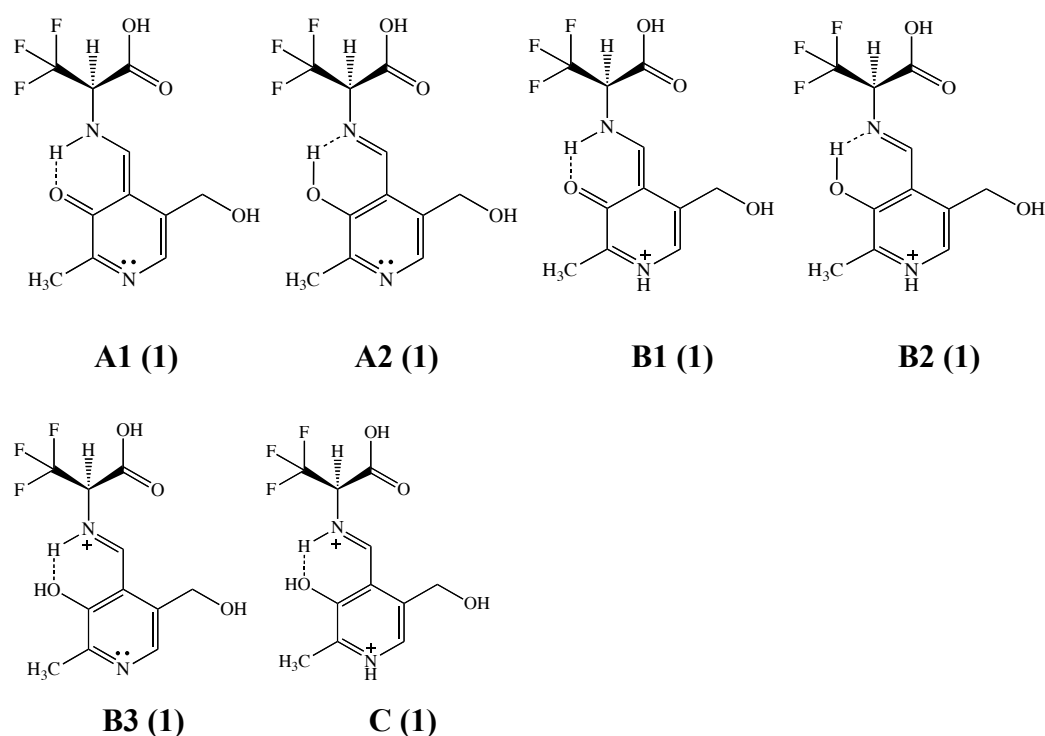


Figure 6-12. External aldimines. Tautomeric forms of the first intermediates of the mechanism of inhibition of alanine racemase, proposed by W. S. Faraci and C. T. Walsh¹ in their possible protonation states depending on the pH

Structures **A1 (1)** and **A2 (1)**, and **B1 (1)**, **B2 (1)** and **B3 (1)** are tautomers. A comparison of their energies relative to each other excluded two of these structures as possible intermediates in the process.

Structure	A1 (1)	A2 (1)		
Relative energy	41 kJ/mol	0 kJ/mol		
Structure	B1 (1)	B2 (1)	B3 (1)	
Relative energy	17kJ/mol	0 kJ/mol	51 kJ/mol	

Table 6-3. Energies of the **A** and **B** relative to the most stable tautomers at the M06-2X/6-311G(2df, p) level in vacuum

The relative energies of the **A1 (1)** and **B3 (1)** were high, so they were eliminated as possible tautomeric structures. The most stable **A (1)** and **B (1)** structures were the enol-imine forms **A2 (1)** and **B2 (1)**. The **B1 (1)** structure was only 17 kJ/mol higher in energy than the most stable structure **B2 (1)**, so it was considered as a possible tautomer.

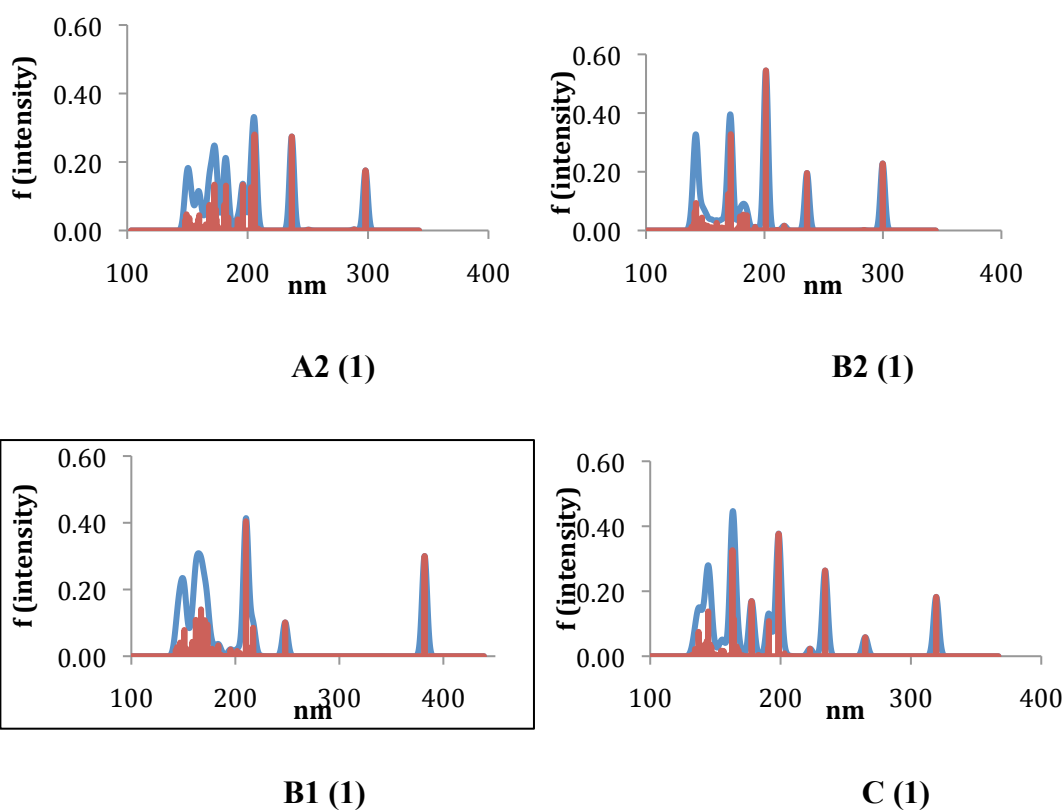


Figure 6-13. Calculated spectra of the external aldimines with M06-2X/6-311+G(2df, p) in IEFPCM continuum water

The calculated spectra of **A2 (1)**, **B2 (1)** and **C (1)** looked like it was expected- with an initial peak in the 300-330nm region of the UV spectrum. In contrast, the first peak in the **B1 (1)** spectrum appears at 382 nm (Figure 6-13). If the average error of 0.63×10^{-19} J for ΔE was applied, it gave a peak with λ_{max} in the visible region of the spectrum with a wavelength $\lambda_{\text{max}} = 434$ nm is given. This would appear as an absorption in the violet region of the visible spectrum. The orbital representations gave a clearer view of the reason why this peak appears at such a long wavelength:

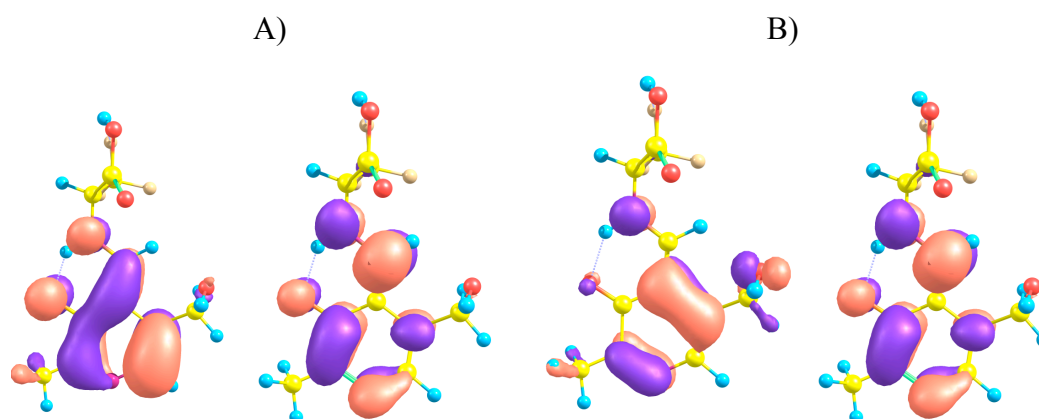


Figure 6-14. Orbitals, generated at the M05-2X/6-311+G(2df, p) level in gas phase, corresponding to the first two peaks in the **B1 (1)** spectrum: A) Excited State 2: $E = 3.3568$ eV, $\lambda = 369.35$ nm, $f = 0.2840$, 75 \rightarrow 76 orbitals transition, $CI = 0.69921$; B) Excited State 4: $E = 4.9073$ eV, $\lambda = 252.65$ nm, $f = 0.1760$, 73 \rightarrow 76 orbitals transition, $CI = 0.68291$

The orbitals for the first peak (second excited state) showed a HOMO-LUMO $\pi \rightarrow \pi^*$ transition. The π orbital was spread over the extended conjugated π system in the carbonyl resonance structure of **B1 (1)**. The second peak corresponding to excited state 4 at 253 nm is again a $\pi \rightarrow \pi^*$ transition from HOMO-2 to LUMO and covers the π system.

The other tautomer **B2 (1)** had a peak in the TDDFT spectrum at 300 nm and with the correction gave an absorption at 331 nm. Both the ketoenamine **B1 (1)** and enol-imine **B2 (1)** would give two absorption peaks at 434 nm and 331 nm. This agrees perfectly with the experimental data.¹⁴⁰ The low and the high pH forms of the Schiff bases **A2 (1)** and **C (1)** had calculated absorption maxima at 298 nm and 319 nm and would

give experimental peaks in the electronic spectrum at 329 nm and 354 nm if we apply the average error correction.

6.3.3 Structures and spectra of quinonoids (1a)

The deprotonation of the external aldimine gave quinonoid structures in protonated pyridine ring species. The only exception was the non-quinonoid spectrum in **A2 (1a)**. After the deprotonation, **B1 (1a)** tautomerised to **B2 (1a)** as a product (Figure 6-15):

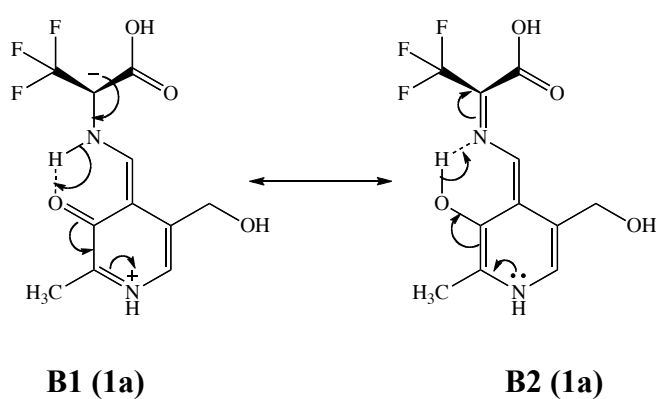


Figure 6-15. Tautomerisation of **B1 (1a)** to quinonoid structure **B2 (1a)** optimised with M06-2X/6-311G(2df, p) in gas phase

The three final **(1a)** structures after deprotonation were:

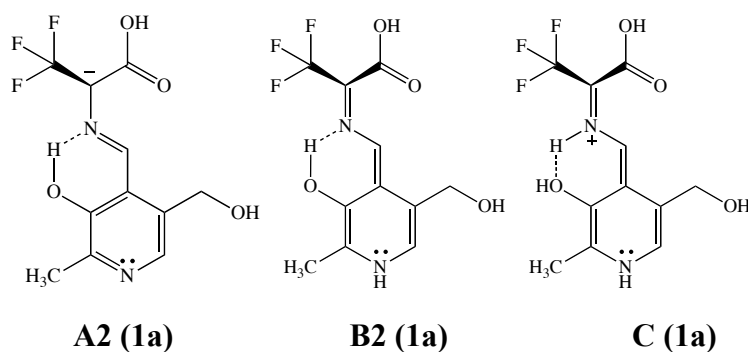


Figure 6-16. Structures of deprotonated external aldimines **(1a)** intermidites (second step in the mechanism of the alanine racemase inhibition)¹

The calculated spectra are presented in Figure 6-17:

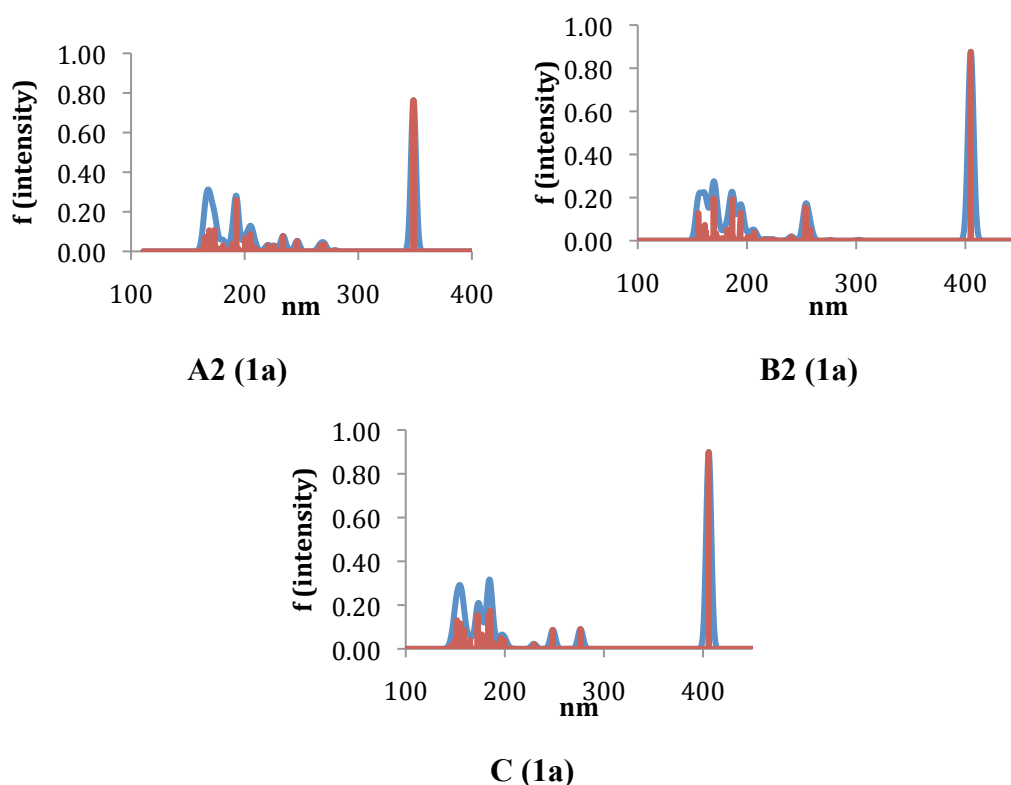


Figure 6-17. Calculated spectra of the **(1a)** compounds at the M06-2X/6-311G(2df, p) in IEFPCM continuum water

A2 (1a) gave a corrected absorption at 392 nm. The two quinonoids **B2 (1a)** and **C (1a)** absorbed at $\lambda_{\text{max}} = 405$ and $\lambda_{\text{max}} = 406$ nm and their corrected absorption was 465 nm (Figure 6-17). Looking at the orbital representation the first excitation was a $\pi \rightarrow \pi^*$ transition (HOMO-LUMO) in the extended conjugated π -system (Figure 6-18).

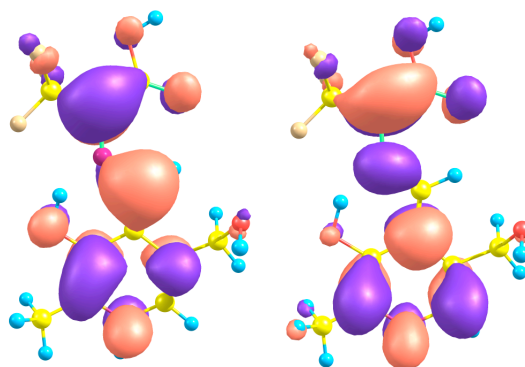


Figure 6-18. Orbital at the M05-2X/6-311+G(2df, p) level in gas phase of the first excited state in the quinonoid form **(1a)**: Excited State1: E= 3.0614 eV, $\lambda = 405$ nm, f= 0.8761, 75 -> 76 orbitals transition, CI= 0.70307

The different protonation states of the **A2** and **B2** aldimines and their quinonoids were connected. The reaction energies showed that they could pass from one to another through a reaction of protonation and deprotonation (Figure 6-19). $\text{CH}_3\text{NH}_3^+/\text{CH}_3\text{NH}_2$ were used as the acid/base pair. All the geometries were optimised in the gas phase with the standard functional and basis set M06-2X/6-311+G(2df, p) and the energies were calculated in water (IEFPCM continuum model).

The quinonoids were more stable than their aldimine forms. The reaction energy for the protonation of **A2** (**1a**) to **B2** (**1a**) was -6 kJ/mol, so it could be considered as a possible reaction. **A2** (**1**) to **B2** (**1a**) are very close in energy (19kJ/mol) so can tautomerise.

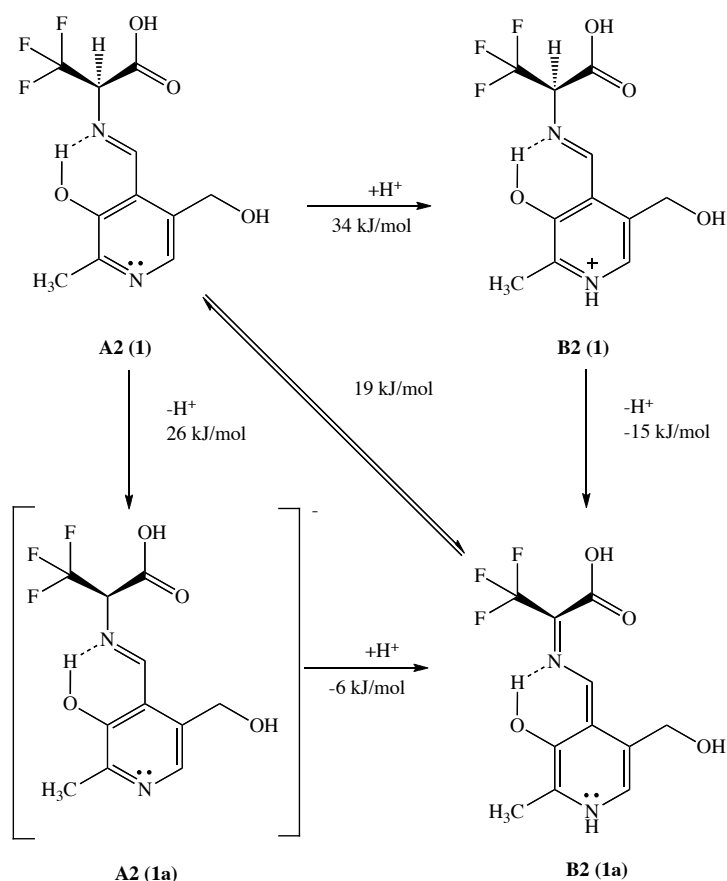


Figure 6-19. Reaction energies of the protonation/deprotonation against $\text{CH}_3\text{NH}_3^+/\text{CH}_3\text{NH}_2$ of aldimines and quinonoids of **A2** and **B2** at the M06-2X/6-311G(2df, p) level in vacuum

6.3.4 Structures and spectra of β -difluoro- α,β -unsaturated imine

After the loss of the fluoride anion the bands of the electronic spectra of the β -difluoro- α,β -unsaturated imines were in the UV range (Figure 6-20). The calculated λ_{max} of the first peaks were 309 nm, 315 nm and 322 nm and the expected experimental values were 345 nm, 350nm and 372 nm (Figure 6-21).

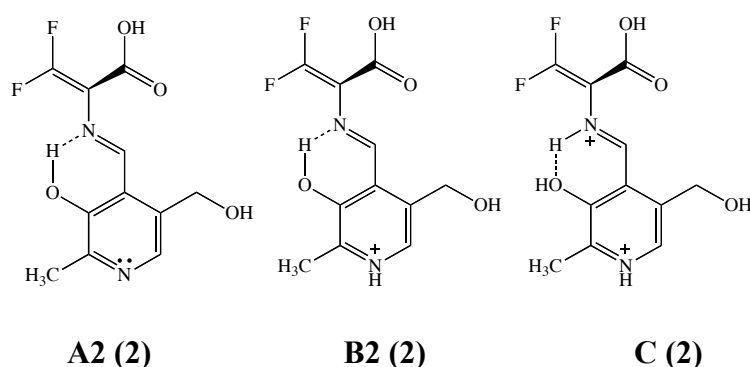


Figure 6-20. Structures of β -difluoro- α,β -unsaturated imine intermediates (compound **2**) optimised with M06-2X/6-311G(2df, p) in vacuum (third step mechanism proposed by Wash *et al.* for the alanine racemase inhibition)¹

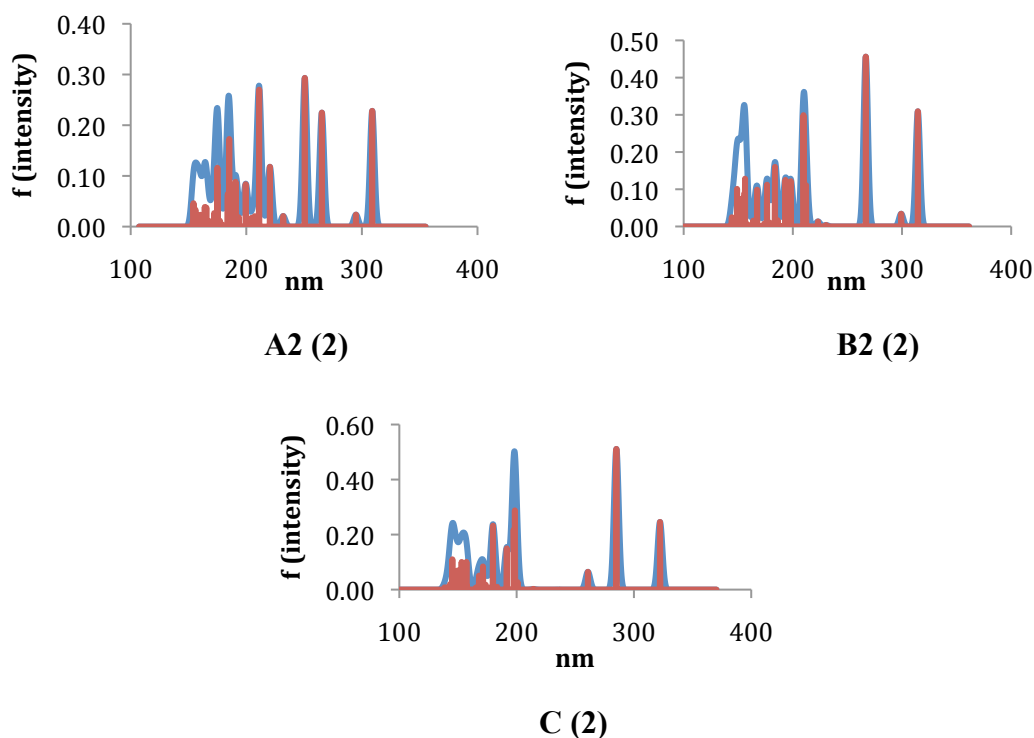


Figure 6-21. Calculated spectra of the β -difluoro- α,β -unsaturated imines in IEFPCM continuum water using M06-2X/6-311G(2df, p)

The total reaction energies of the elimination of hydrofluoric acid (HF) from the aldimine (**1**) to compound (**2**) were relatively high: 108 kJ/mol for **A2**, 102 kJ/mol for **B2** and 98 kJ/mol for **C**.

6.3.5 Structures and spectra of the enzyme-pyridoxal aldimine complexes

The unsaturated imine complex of pyridoxal and the inhibitor is subject to a nucleophilic attack of group “:X” from a residue in the enzyme. Usually this happens to be a lysine residue (Lys38 in *B. stearothermophilus*).¹ For the computational studies a model of the lysine amino acid residue- methylamine was used. The products of the reactions and their spectra were as follows:

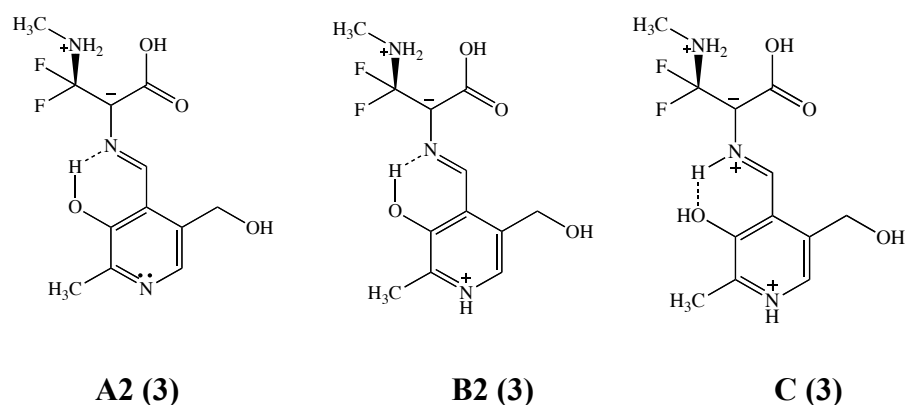
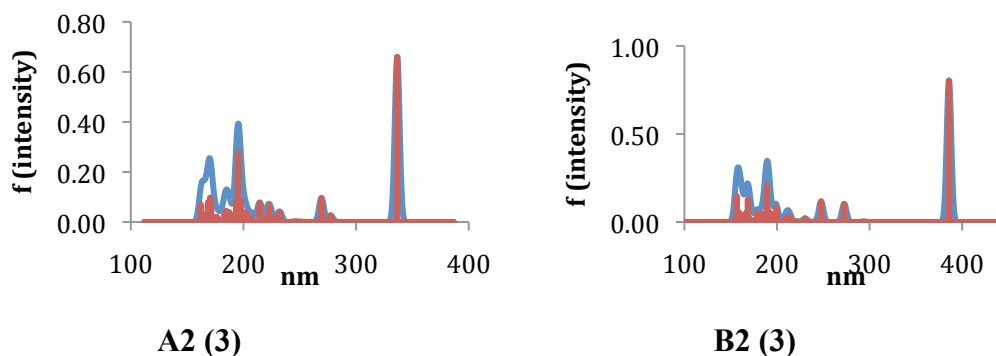
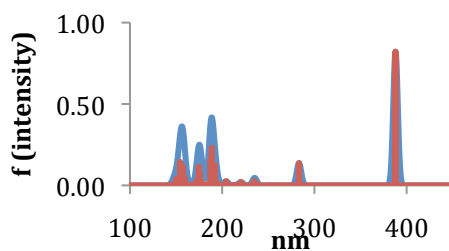


Figure 6-22. Addition products from the reaction between the enzyme residue and the pyridoxal-inhibitor complex in the Walsh *et al.* mechanism¹





C (3)

Figure 6-23. Calculated spectra of the **(3)** structures in different protonation states in IEFPCM continuum water using M06-2X/6-311G(2df, p)

The negative charge on the C_α -atom was delocalised over the whole molecule and this caused a shift of the absorption band to a longer wavelength (Figure 6-23). The Mulliken atomic charges for the structures in two different protonation states in the pyridine ring are shown in Figure 6-24. The comparison of the charges before and after the nucleophilic attack of the enzyme residue showed a delocalisation of the negative charge over the whole structure.

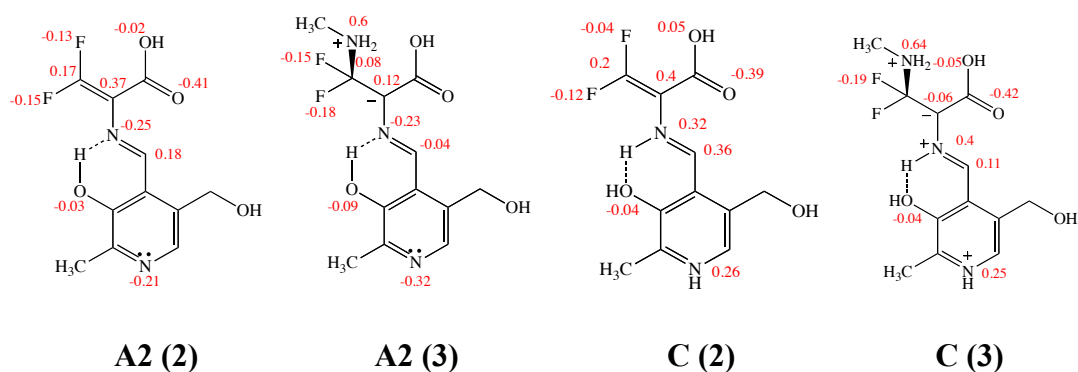


Figure 6-24. Mulliken atomic charges of the structures **A2 (3)** and **C (3)** before and after a nucleophilic attack

The electronegative atoms in proximity to nitrogen and fluorine destabilise the negative charge. The structures with delocalised negative charge and protonated nitrogen of the pyridine ring had, in this case, quinonoid as resonance structures (Figure 6-25). The bond distances in the B2 (3) structure proved the existence of the stable quinonoid structure (Figure 6-26).

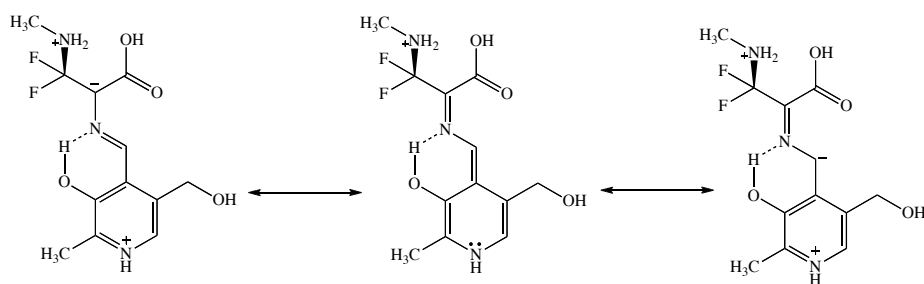


Figure 6-25. Resonance structures of **B2 (3)**

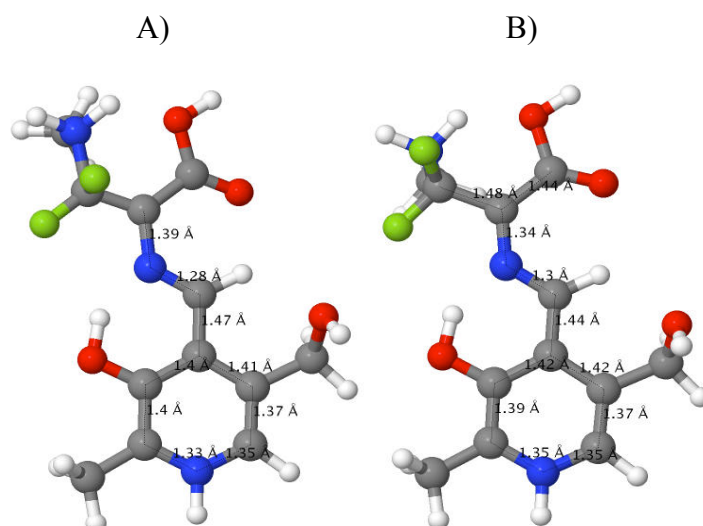


Figure 6-26. Bond distances in **B2 (3)** before (A) and after the optimisation (B) at the M06-2X/6-311G(2df, p) level in gas phase

The longest wavelength peaks of **A2 (3)**, **B2 (3)** and **C (3)** appeared at 336 nm, 386 nm and 387 nm. After correction these values became 377 nm, 440 nm and 441 nm.

6.3.6 Structures and spectra of (3a)

The deprotonation of the compound **(3)** gave quinonoid structures in **B2 (3a)** and **C (3a)** and a structure with delocalised charge for **A2 (3a)** (Figure 6-27). The quinonoids' first peaks appeared at 413 nm and 418 nm, and they were corrected to 475 nm and 482 nm. The **A2 (3a)** had a peak at $\lambda_{\text{max}} = 366$ nm, and would give a 415 nm band in the violet Vis-spectrum (Figure 6-28).

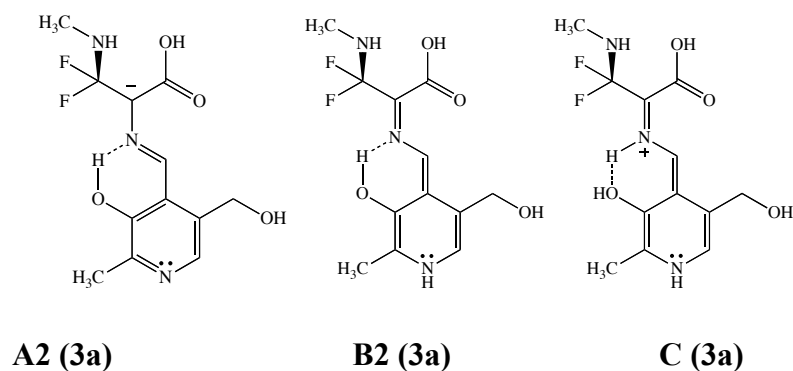


Figure 6-27. Structures of **(3a)** compounds, deprotonated intermediates with **B2 (3a)** and **C (3a)** quinonoid structures

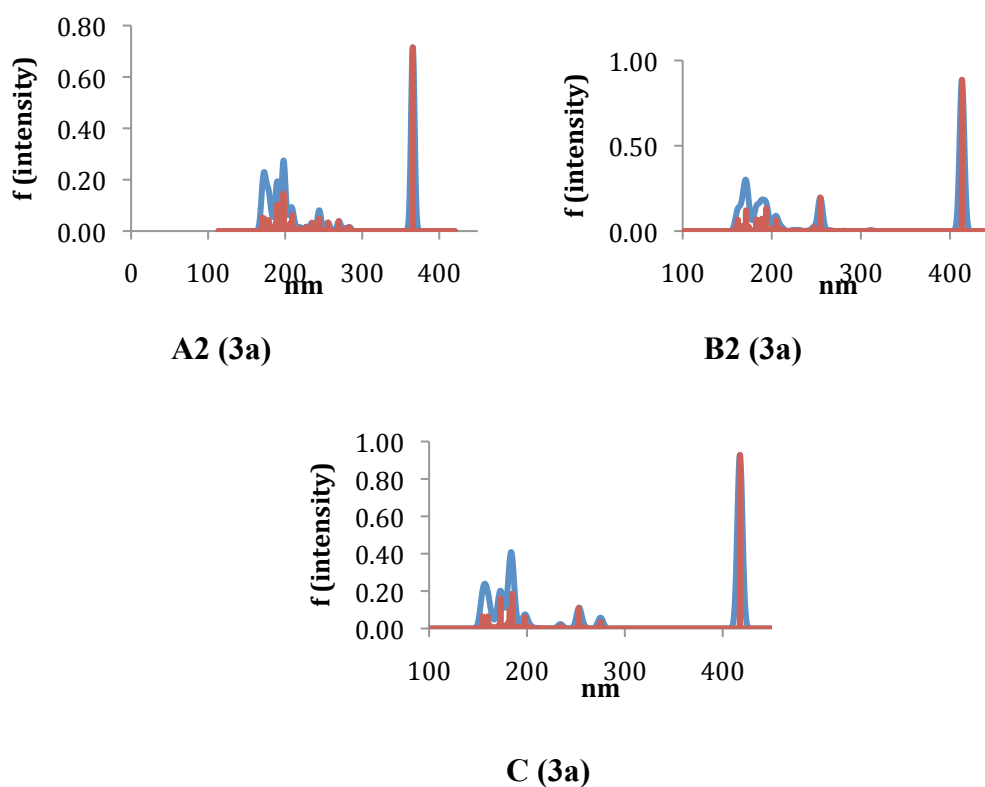


Figure 6-28. Calculated spectra of **(3a)** species with M06-2X/6-311G(2df, p) in IEFPCM continuum water

The reaction energies for the deprotonation of the nucleophilic amino group against $\text{CH}_3\text{NH}_3^+/\text{CH}_3\text{NH}_2$ were -31 kJ/mol, -38 kJ/mol and -71 kJ/mol for the **A2 (3a)**, **B2 (3a)** and **C (3a)** compounds.

6.3.7 Structures and spectra of the inactive complex (4)

In the inactive complex **(4)** the inhibitor-pyridoxal complex was covalently-bound to the enzyme nucleophilic group, in this case the amino group (Figure 6-29). The species with a protonated pyridine ring had resonance structures where the lone pair of the amino group was donated and extended conjugated systems were formed (Figure 6-30).

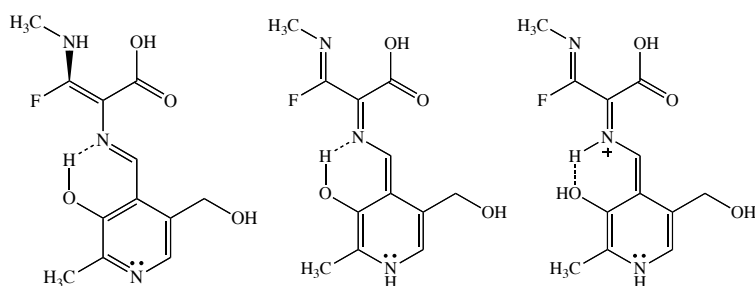


Figure 6-29. Final products **(4)** of the mechanism of inhibition of the alanine racamase by β,β,β -trifluoroalanine¹

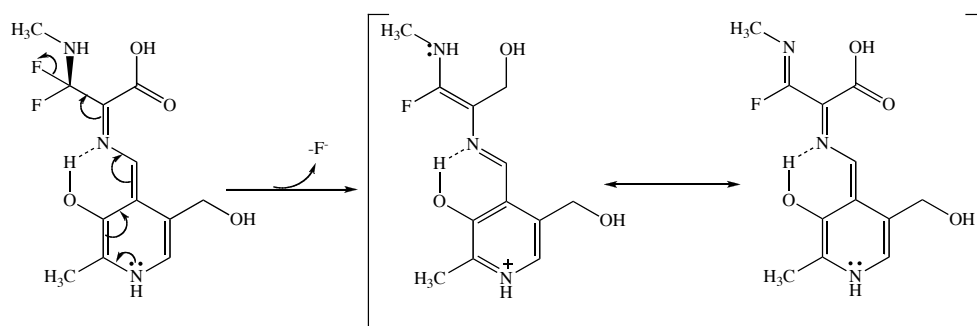


Figure 6-30. Resonance forms of **B2 (4)** product

The spectra of the final products gave λ_{\max} of 328 nm, 361nm and 378 nm. With the calculated shift values the predicted bands shifted to 366nm, 408 nm and 430 nm for **A2 (4)**, **B2 (4)** and **C (4)**. The last two compounds would absorb in the violet region of the visible spectrum.

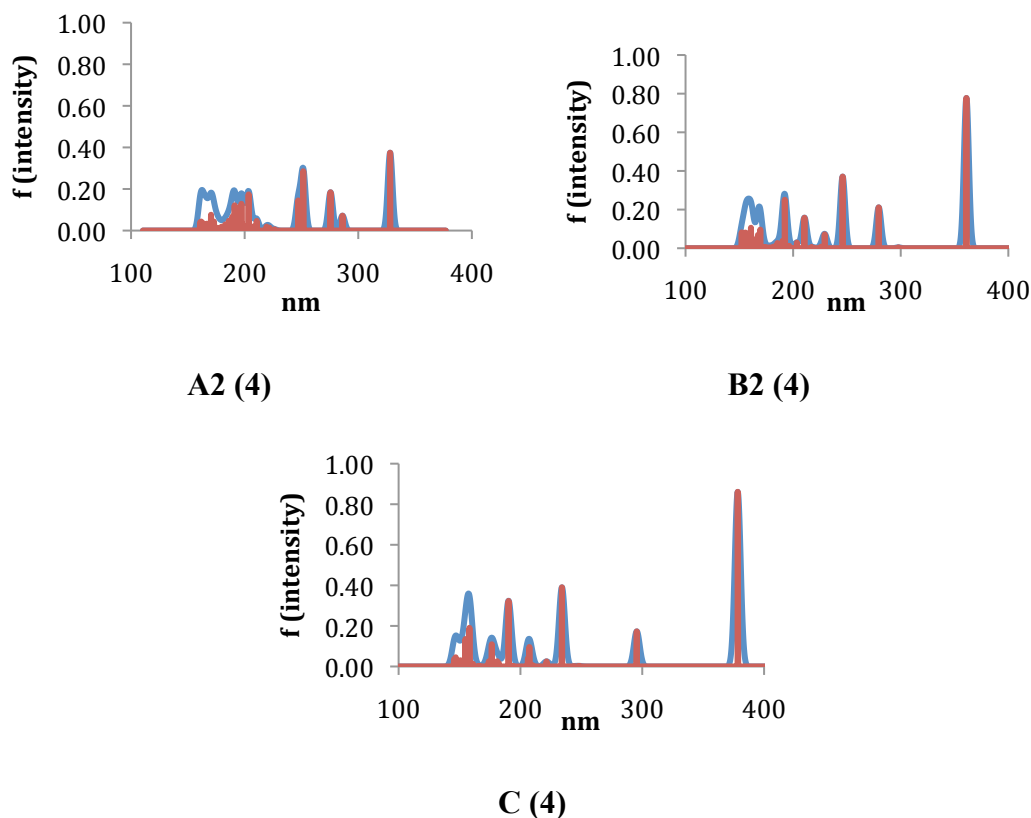


Figure 6-31. Calculated spectra of **(4)** compounds at the M06-2X/6-311G(2df, p) level in IEFPCM continuum water

6.4 Conclusions

PLP (pyridoxal 5'-phosphate) is a co-factor in various enzymes and in particular in alanine racemase, an enzyme which catalyses the racemisation of L-alanine to D-alanine. The inhibitor β,β,β -trifluoroalanine blocks the activity of this enzyme reacting with the PLP co-factor and producing an external aldimine. The co-factor derivatives implied in the alanine racemase inhibition mechanism were optimised and electronic spectra of the reactants, intermediates and products were calculated.

Pyridoxal 5'-phosphate and pyridoxal were used for validation of the spectra, calculated using TDDFT after a comparison with the experimental data. The structures were optimised and spectra were calculated in the gas phase and IEFPCM water with M06-2X/6-311G(2df, p). The calculated spectra were compared visually to the experimental ones. Excitation energies of the calculated and experimental spectra were also compared and spectra were calculated in the gas phase and continuum

water. The conclusion was that the spectra had to be calculated using water as a solvent but the optimisation could be performed in the gas phase. The method used for the UV/Visible spectra was M06-2X/6-311G(2df, p) in IEFPCM water. The phosphate group introduced artificial charge-transfer transitions. It is not active in the relevant part of the spectra and was replaced by a hydroxyl group.

The Schiff base of the ketoenamine gave a band in the violet region of the visible spectrum with λ_{max} of 434 nm. The keto forms were more unstable than their enol imine tautomers. The **B1** tautomer was less stable with only 17 kJ/mol in the first step of the mechanism but in the next step it tautomerised to the more stable **B2** form.

The protonated quinonoid forms of **1a** and **3a** were presented as visible-colour compounds in the blue region of the spectrum with wavelengths from 465 to 482 nm. The non-protonated **1a** had a band on the border between the UV and visible regions (392 nm), which switched to a longer wavelength in **3a** (415 nm). The formations of the quinonoid forms were energetically favourable steps along the mechanism.

The β -difluoro- α,β -unsaturated imine had peaks entirely in the UV-part of the spectrum.

The model of the enzyme-pyridoxal complex (**3**) and the final inhibited product (**4**) for the protonated forms had peaks in the violet visible region, compared to the non-protonated forms, which appeared in the UV-region.

In conclusion, the protonated pyridoxal compounds at a lower pH gave spectra with bands in the visible region. These structures had resonance quinonoid forms with extended conjugated systems compared to the non-protonated pyridoxals, which stayed in the UV-region.

The coloured compounds absorb light with specific wavelength from white light; the light which it is not absorbed passes through unaffected. The primary colours in the colour wheel have a complementary colour and the mixtures of each of the primary colours with their complementary colour give white light. When one of the two colours is absorbed from a coloured compound, the other passes through and gives the colour of the compound.

The investigated compounds in the mechanism of the inhibition of the alanine racemase would then be yellow if they absorb in the violet region or orange if the band is in the blue region of the visible spectrum.

The calculated spectra and relative energies for all the steps and tautomers in their possible protonation states of the mechanism of inhibition of the alanine racemase were presented in this chapter and the results are ready to be compared to experimental data.

Chapter 7 Theoretical and X-ray crystallographic evidence of a fluorine-imine gauche effect: an addendum to Dunathan's hypothesis

7.1 Introduction

Dunathan's stereoelectronic hypothesis to correlate the conformation and stereospecificity of pyridoxal phosphate (PLP **1**) dependent enzymatic processes remains a milestone discovery in rationalising the mechanistic intricacies of vitamin B₆ enzymes.¹⁴¹ Of principal importance is the notion that σ -bonds may be activated by an adjacent π -system (Figure 7-1). Consequently, the co-factor-derived Schiff base (**2**) that is central to a plethora of enzyme mechanisms, including those of decarboxylases and racemases, shares a common topological transition state dependence on the σ -bond being aligned with the π -system of the pyridoxal imine: in essence, maximum σ - π overlap regulates reaction specificity.^{129,142-145}

Unsurprisingly, this mechanistic understanding has led to the design of numerous small molecule inhibitors for the treatment of neurodegenerative diseases as well as PET imaging agents.^{1,146-149} Many of these pharmaceuticals are fluorinated at the β -position such that elimination from the transient quinonoid intermediate (**3**) generates an electron sink (**4**), which can covalently modify the enzyme active site, thus rendering it inactive. Pertinent examples of this design approach include α -fluoromethyl-DOPA (**5**), α -difluoromethyl-DOPA and related medicinally important compounds.^{150,151} However, despite the wealth of literature pertaining to the reactivity and conformational dynamics of pyridoxal-derived Schiff bases,¹⁵²⁻¹⁵⁴ including β -fluoroimines,¹⁵⁵⁻¹⁵⁸ an important conformational issue of this molecular class remains unaddressed: do β -fluoroimines preferentially adopt a *gauche* conformation in a manner consistent with other fluorinated compounds containing a electron-withdrawing group in the β -position.¹⁵⁹⁻¹⁶⁸

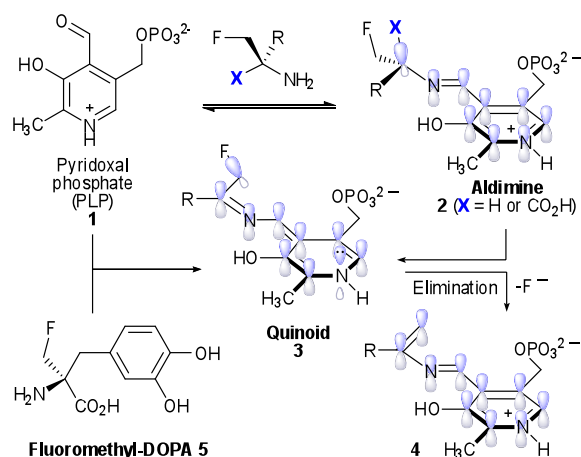


Figure 7-1. The Dunathan Hypothesis to correlate reaction specificity and conformation in PLP-dependent enzymes¹⁴¹

7.1.1 Experimental structures in the solid state

As a starting point for this study, β -fluoroimines **6–8** were evaluated by single crystal X-ray diffraction analysis (Table 7-1).¹⁶⁹ In the case of the imine **6** (entry 1), in which there are no dominant steric interactions, the interatomic distances and angles are within the expected range, and a clear *gauche* conformation is observed between the fluorine and nitrogen centres [$\varphi_{\text{N4-C1-C2-F3}} = \pm 70.0^\circ$]. Interestingly, the planar section of the β -fluoroimine **6** forms a π - π interaction at a distance of 3.4 Å with a neighbouring molecule translated along the *c* axis.

In the crystal structure **7** (entry 2), fluorine adopts a *syn*-clinal orientation relative to the imine nitrogen [$\varphi_{\text{NCCF}} = -74.2^\circ$]. The nitrobenzyl group is rotated out of the imine plane [$\varphi_{\text{N18-C19-C20-C21}} = -25.5^\circ$], likely induced by crystal packing forces where π - π stacking with an adjacent nitrobenzyl moiety (distance of the two ring centroids = 3.75 Å) is the most dominant interaction.

Finally, the attention was focussed on **8** (entry 3); a species that is structurally analogous to the transient intermediates implicated in the disruption of many PLP-dependent enzymatic processes. This aldimine (**8**) has a racemic and an enantiomerically pure form. Importantly, a clear *gauche* effect is observed [$\varphi_{\text{NCCF}} = \pm 67^\circ$]. The orthogonal orientation of the (normally phosphorylated) primary hydroxyl group relative to the pyridine ring is also noteworthy, and all chemically intuitive hydrogen bonding patterns are satisfied.

It is important to note that in all cases the (*E*)-configured imines were isolated exclusively, thus minimising steric stress, and that the non-sterically congested systems (**6** and **8**) adopt conformations in which the nitrogen lone pair points away from the fluorine atom to minimize electronic repulsion. Whereas the solid state conformations of **6** and **8** clearly show opposing C–F and C–N dipoles, this observation does not hold true for β -fluoroimine **7**, in which the nitrogen lone pair and the C–F bond are parallel, thus minimising 1,3-allylic strain (Figure 7-2).

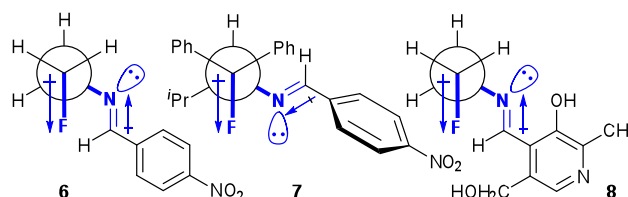


Figure 7-2 Newman projections of β -fluoroimines **6**, **7** and **8**

	β -fluoroimine	X-ray Structure	Torsion angle
1			$\phi_{\text{NCCF}} = \pm 70.0^a$
2			$\phi_{\text{NCCF}} = -74.2^b$
3			$\phi_{\text{NCCF}} = \pm 67.0 / \pm 66.9^a$ $\phi_{\text{NCCF}} = 67.0 / -66.3^c$

[a] Non-polar space group; both gauche conformers observed; two symmetry-independent molecules in the unit cell. [b] Polar space group; configuration of the starting material known. [c] Polar space group

Table 7-1. Selected X-ray data for fluoroimines **6**, **7**, and **8**, produced by Gilmour and co-workers¹⁶⁹

7.2 Computational methodology

All calculations were done with *Gaussian 09*¹³¹ using density-functional theory. The M06-2X hybrid meta-GGA exchange-correlation functional was used throughout, which has been shown to yield superior accuracy not only for main-group thermochemistry, but in particular for non-covalent interactions, including dispersion and hydrogen bonding.¹³² Even though such interactions are not expected to be dominant in the systems studied here, they may well influence conformational preference. We used standard Pople-style basis sets, starting with 6-31+G(d, p), where the addition of diffuse functions on non-hydrogen atoms enormously improves energetics when using DFT methods, comparable to what is achieved with triple- ζ basis sets.¹⁷⁰ Minima were re-optimized using 6-311G(2df, p) and spot-checked with 6-311+G(2df, p). Except for (*E*)-**8c**, the effect of basis set on relative energies was <1 kJ mol⁻¹. Calculations in water solvent used the IEFPCM polarizable continuum model with UFF atomic radii for the cavity construction. Non-electrostatic contributions to solvation were not included (which corresponds to the default solvation treatment in *Gaussian 09*). Default convergence criteria were applied for the SCF and in geometry optimizations. All stationary points were confirmed as minima by a positive-definite Hessian. Values for rotational barriers were estimated from the maxima of the torsion profiles.

7.3 Results and Discussion

7.3.1 Computational conformational analysis

7.3.1.1 Conformational preferences of β -fluoroimines

In an attempt to quantify the apparent *gauche* conformational preference of the β -fluoroethylimines described in Table 7-2, it was embarked upon a theoretical study of related systems at the DFT level. Initially, the NCCF bond rotational profile of the imine derived from β -fluoroethylamine and acetaldehyde was calculated in vacuum and compared with the corresponding de-halogenated parent system (Figure 7-3;

Table 7-2, entries 1–6). The parent system shows the usual symmetric rotation profile with three equivalent minima separated by barriers of $\sim 13 \text{ kJ mol}^{-1}$.

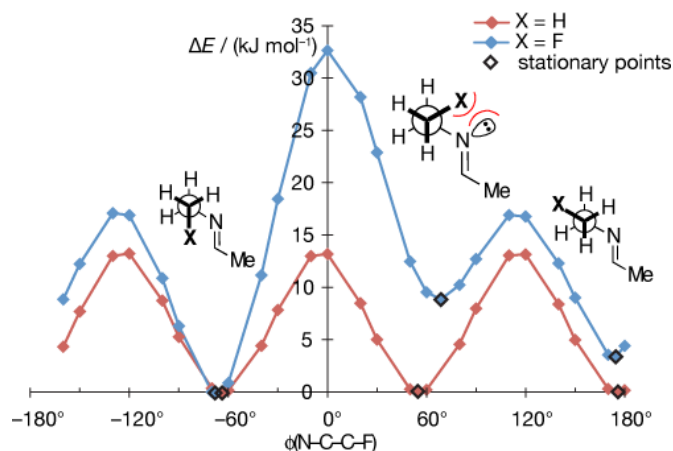


Figure 7-3. Bond rotation profiles of acetaldehyde-derived imines (F vs. H), calculated in vacuum at the M06-2X/6-311G(2df, p) level

Entry	X		Conformer	$\Delta E / (\text{kJ/mol})$	$\phi (\text{NCCX}) / ^\circ$
1	H	Me	<i>−gauche</i>	0	−64
2	H	Me	<i>+gauche</i>	0	55
3	H	Me	<i>anti</i>	0	176
4	F	Me	<i>−gauche</i>	0	−68
5	F	Me	<i>+gauche</i>	9	69
6	F	Me	<i>anti</i>	4	176
7	F	H	<i>−gauche</i>	0	−68
8	F	H	<i>+gauche</i>	8	69
9	F	H	<i>anti</i>	4	175
10	F	4-NO ₂ Ph(6)	<i>−gauche</i>	0	−67
11	F	4-NO ₂ Ph(6)	<i>+gauche</i>	8	68
12	F	4-NO ₂ Ph(6)	<i>anti</i>	6	176
13	F	4-MeOPh	<i>−gauche</i>	0	−68
14	F	4-MeOPh	<i>+gauche</i>	8	68
15	F	4-MeOPh	<i>anti</i>	3	176

Table 7-2. Calculated relative conformational energies and torsion angles of ethylimines derived from selected aldehydes RCOH. The results were obtained with M06-2X/6-311G(2df, p) in vacuum

In the β -fluoro system, the symmetry is broken, with the *−gauche* conformer being

the most stable, followed by *anti* (+4 kJ mol⁻¹ relative to *-gauche*) and *+gauche* (+9 kJ mol⁻¹). The preference for *-gauche* over *anti* is explained by a hyperconjugative $\sigma_{\text{C-H}} \rightarrow \sigma^*_{\text{C-F}}$ interaction. To maximize the stabilizing overlap between the strongly electron-accepting $\sigma^*_{\text{C-F}}$ orbital and the donating vicinal $\sigma_{\text{C-H}}$ orbital (which is a better donor than $\sigma_{\text{C-N}}$), the C-F bond is oriented *anti* to either vicinal C-H bond. Hence results a stereoelectronic preference for C-F being *gauche* to C-N. This deviation from simple steric and electrostatic arguments is known as the stereoelectronic “*gauche* effect”, the prototype of which is 1,2- difluoroethane.¹⁷¹ However, the *gauche* effect does not differentiate between the two possible *gauche* conformers. The observed destabilisation of the *+gauche* conformer in the β -fluoroimines investigated here is caused by the overriding electrostatic repulsion between the imine and fluorine lone pairs, as schematically indicated in Figure 7-1. Qualitatively and quantitatively similar behaviour was previously reported for other systems where the fluorine is in the β -position with respect to a lone-pair bearing heteroatom, such as β -fluoroamines or β -fluoroalcohols.^{159,162} The preferred *-gauche* minimum is protected by barriers of ~33 kJ mol⁻¹ (*-gauche* \rightarrow *+gauche*) and ~17 kJ mol⁻¹ (*-gauche* \rightarrow *anti*); the *anti* \rightarrow *+gauche* barrier is ~13 kJ mol⁻¹. These energetic data suggest that all three conformers are accessible in equilibrium at 298 K, with relative populations of 1 : 0.2 : 0.03 (*-gauche* : *anti* : *+gauche*).

Based on the electronic origin of the preference for *-gauche* over *anti*, one expects that there should be a dependence on the electronic nature of the imine: The *anti* conformer should be the more disfavoured the more electron-withdrawing the imine. To test this hypothesis, we investigated related β -fluoroimines derived from selected aldehydes RCOH (R = H, 4-NO₂Ph (**6**), 4-MeOPh); see Table 7-2, entries 7–15. The comparison shows that there is only a small effect of the R group on the NCCF conformational preferences. The stabilisation of *-gauche* over *anti* is slightly more pronounced (6 vs. 3 kJ mol⁻¹) for the electron-deficient aromatic imine (R = 4-NO₂Ph, entries 10–12) as compared to the electron-rich one (R = 4-MeOPh, entries 13–15). Albeit small, this effect is consistent with the stereoelectronic explanation of the *gauche* effect, whereby the *anti* arrangement of acceptor bonds is disfavoured. In all cases, the order of stability remains the same: *-gauche* < *anti* < *+gauche*.

7.3.1.2 Natural Bond Orbital analyses

The extent of the stereoelectronic *gauche* effect can be quantified by considering the stabilization gained from the hyperconjugative donor–acceptor interaction. Within the framework of the Natural Bond Orbital (NBO) approach,¹⁷² the interaction energy due to the delocalization of electron density from a donor NBO to an acceptor NBO can be calculated from a second-order perturbation theory expression. NBOs are localized orbitals constructed such as to provide the most accurate Lewis-like bonding picture, based on a given *N*-electron density, which can be obtained from any electronic-structure method. Specifically for the case of the NCCF torsion, the NBO method yields localized bonding and anti-bonding NBOs for the C–F, C–H, and C–N bonds, and it quantifies the energy due to any donor–acceptor interactions between them.

In agreement with previous studies¹⁶² we found that the total donor–acceptor stabilization within the NCH₂CH₂F fragment is dominated by interactions between *anti*-oriented vicinal bonds. The *gauche* vicinal interactions are minor and change little upon rotation; the geminal interactions, although significant, are unaffected by rotation. We therefore used the sum of the six vicinal *anti* donor–acceptor interaction energies as a measure for the hyperconjugative stabilization, E_{deloc} . For instance, for the three conformers of the 4-nitrobenzalimine **6**, E_{deloc} amounts to –66, –64, and –57 kJ mol^{–1} for the *–gauche*, *+gauche*, and *anti* conformer, respectively. In the *gauche* conformers, the single largest contributor is indeed the $\sigma_{\text{C–H}} \rightarrow \sigma^*_{\text{C–F}}$ interaction with about –20 kJ mol^{–1}. In **6**, the stereoelectronic preference for *–gauche* over *anti* is therefore $\Delta E_{\text{deloc}}(-g, a) = -8.5 \text{ kJ mol}^{-1}$.

For the 4-methoxybenzalimine, we expected less stereoelectronic *gauche* stabilization, in line with the slightly smaller *gauche* preference, because the imine is less electron- withdrawing, as discussed above. However, $\Delta E_{\text{deloc}}(-g, a) = -8.2 \text{ kJ mol}^{-1}$ for the 4-methoxybenzalimine, insignificantly different from **6**, which hence cannot account for the difference in *gauche* preference. We next inspected the atomic partial charges derived from the NBO procedure. They are essentially independent of the torsional conformation. However, in the *gauche* conformers, the two electronegative atoms F and N, which also bear negative NBO charges, are relatively close to one another, about 2.85 Å. On electrostatic grounds, the *gauche* conformers are thus destabilized relative to the *anti* conformer. Whereas the NBO charge on fluorine is the same in both benzalimines ($q_{\text{F}} = -0.40e$), the imine nitrogen is slightly

more negatively charged in the electron-rich 4-methoxy derivative ($q_N = -0.48e$) than in the 4-nitro derivative **6** ($q_N = -0.45e$). This amounts to a difference in electrostatic repulsion of about 5 kJ mol^{-1} (also considering the small change in the F–N distance, which is 0.02 \AA shorter in **6**). The larger *gauche* preference in **6** compared to the 4-methoxybenzalimine can thus be explained by a smaller electrostatic destabilization of the *gauche* conformer.

The conclusion from this part is that the electronic nature of the imine (electron-rich vs. electron-deficient) does have an effect, albeit small, on the NCCF torsional preference of β -fluoroimines. The *gauche* effect is slightly more pronounced in electron-poor imines. The reason for this is, however, not a larger hyperconjugative stabilization, but the reduced electrostatic repulsion between the two electronegative atoms F and N.

7.3.1.3 Conformational preferences of β -fluoropyridoximine

For the pyridoximine **8**, it was considered first of all the relative stabilities of the possible tautomeric and protonation forms with overall neutral charge, **8a–d** (Table 7-3). In both vacuum and water, the canonical enol-imine form **8a** is preferred. This agrees with the X-ray structure **8**, which was unambiguously identified with the enol-imine form **8a** by comparing computed with experimental structural parameters. The keto-enamine tautomer **8b** is energetically competitive in water, where its equilibrium concentration amounts to $\sim 0.5\%$ at 298 K. The quinoid tautomers **8c** and the zwitterion **8d**, however, are only present in vanishingly small amounts under equilibrium conditions in either medium. As expected, the zwitterion is strongly stabilised in the polar solvent as compared to vacuum.

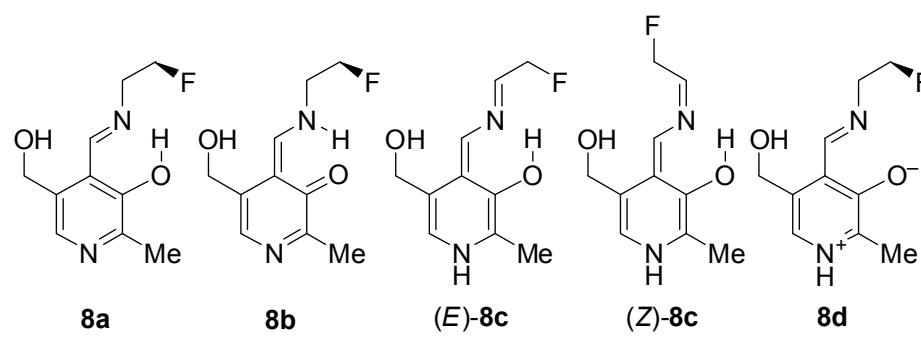
						
8a	8b	(E)-8c	(Z)-8c	8d		
	(-g)-8a	(-g)-8b	(E)-8c	(Z)-8c	(-g)-8d	
$\Delta E /$	vacuum	0	27	76 (-ac)	79 (anti)	115
(kJ mol ⁻¹)	water	0	13	67 (syn)	71 (anti)	65

Table 7-3. Calculated relative stabilities of tautomeric and protonation forms of neutral pyridoximine **8** in vacuum and water. The relative energies of the most stable conformer of each form at the M06-2X/6-311+G(2df, p) level in vacuum and continuum water respectively

For the low-energy tautomers **8a**, **b** as well as for the mechanistically important quinoid **8c**, we calculated NCCF bond rotation profiles in water (Figure 7-4 and Figure 7-5) and optimised the torsional conformers (in both vacuum and water; Table 7-4). Focussing on **8a**, **b** (Figure 7-4; Table 7-4, entries 1–6), we first note that their torsional preferences are essentially identical. Comparing **8a**, **b** in vacuum to the β -fluoroacetalimine (Figure 7-3; Table 7-2, entries 4–6) and the β -fluorobenzaldimines (Table 7-2, entries 10–15) discussed above, the most obvious difference is the stabilization of the +*gauche* conformer, making the *anti* conformer now the least stable one. The order of stability for β -fluoropyridoximine is therefore: $-gauche < +gauche < anti$. The energy of the *anti* relative to $-gauche$ in **8a**, **b** is, however, the same as in the electron-poor imine **6**. The stabilization (or, rather, the smaller destabilization) of the +*gauche* in **8a**, **b** is readily understood when one considers that the nitrogen lone pair is now accepting a hydrogen bond (**8a**) or is part of the conjugated system (**8b**), respectively. The repulsion with the fluorine lone pairs, responsible for the +*gauche* destabilization, is therefore significantly reduced. Otherwise, **8a**, **b** show qualitatively and quantitatively very similar conformational preferences as the previous β -fluoroimines. For instance, the NBO-derived stereoelectronic preference in favour of $-gauche$ is $\Delta E_{\text{deloc}}(-g, a) = -8.7 \text{ kJ mol}^{-1}$ for **8a**, compared to -8.5 kJ mol^{-1} for **6**.

The net effect of solvation in **8a**, **b** is to stabilize the *+gauche* conformer and destabilize the *anti* relative to *−gauche*, however without changing the order of stability. The *anti* destabilization is in agreement with an enhanced stereoelectronic preference; $\Delta E_{\text{deloc}}(-g, a) = -10.1 \text{ kJ mol}^{-1}$ for **8a** in solution, compared to -8.7 kJ mol^{-1} in vacuum. The stabilization of the *+gauche* conformer in **8a** results from a stronger hydrogen bond in solution; the $\text{OH}\cdots\text{N}$ distance reduces by 0.03–0.05 Å upon solvation, which further mitigates the lone pair repulsion.

Entry	Form	Conformer	ΔE / (kJ/mol)	ϕ (NCCX)/ °
1	8a	<i>−gauche</i>	0 (0)	−66 (65)
2	8a	<i>+gauche</i>	6 (3)	64 (62)
3	8a	<i>anti</i>	6 (7)	176 (176)
4	8b	<i>−gauche</i>	0 (0)	−63 (−61)
5	8b	<i>+gauche</i>	4 (2)	62 (60)
6	8b	<i>anti</i>	6 (8)	176 (176)
7	(<i>E</i>)- 8c	<i>−anticlinal</i>	0 (2)	−133 (−122)
8	(<i>E</i>)- 8c	<i>+anticlinal</i>	0 (2)	135 (124)
9	(<i>E</i>)- 8c	<i>syn</i>	2 (0)	0 (0)
10	(<i>Z</i>)- 8c	<i>−gauche</i>	9 (6)	−62 (−62)
11	(<i>Z</i>)- 8c	<i>+gauche</i>	9 (8)	61 (61)
12	(<i>Z</i>)- 8c	<i>anti</i>	3 (5)	179 (178)

Table 7-4. Calculated relative conformational energies of different forms of β -fluoropyridoximine **8**. The results were obtained with M06-2X/6-311G(2df, p) in vacuum and in continuum water (values in parentheses). Energies are relative to the most stable conformer of the respective form in vacuum or water, respectively

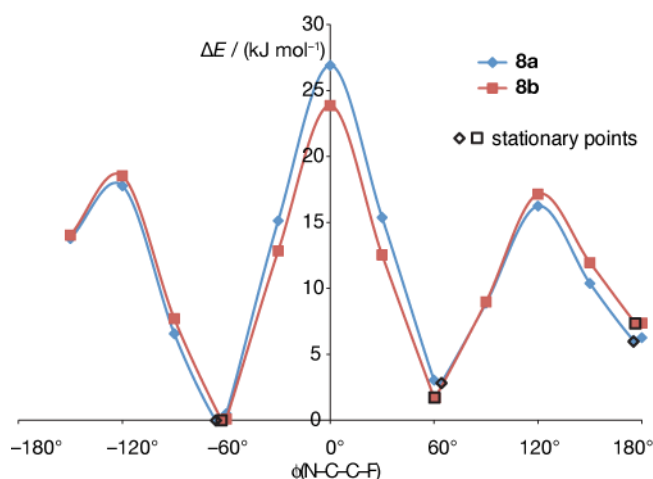


Figure 7-4. Bond rotation profiles of β -fluoropyridoximine tautomers **8a** and **8b**, calculated in water at the M06-2X/6-31+G(d, p) level. Relative energies of stationary points using the larger 6-311G(2df, p) basis set differ by $<1 \text{ kJ mol}^{-1}$ (data not shown).

The quinoid tautomer **8c** has two stereoisomers with respect to the configuration about the external ($\text{N}=\text{C}_\alpha$) imine double bond. Depending on which of the two C_α protons in **8a** is abstracted in the tautomerization, either (*E*)- or (*Z*)-**8c** is formed. The most stable conformer of the *E*-isomer is always preferred (by about 4 kJ mol^{-1}) to the most stable conformer of the *Z*-isomer (Table 7-4, entries 7–12; Figure 7-5). It is important to note that the central bond of the NCCF torsion in **8c** is $\text{sp}^2\text{--sp}^3$, rather than $\text{sp}^3\text{--sp}^3$ as in all the previous cases. This creates a different steric and stereoelectronic situation, and it can therefore be expected torsional preferences to differ from the previously discussed systems.

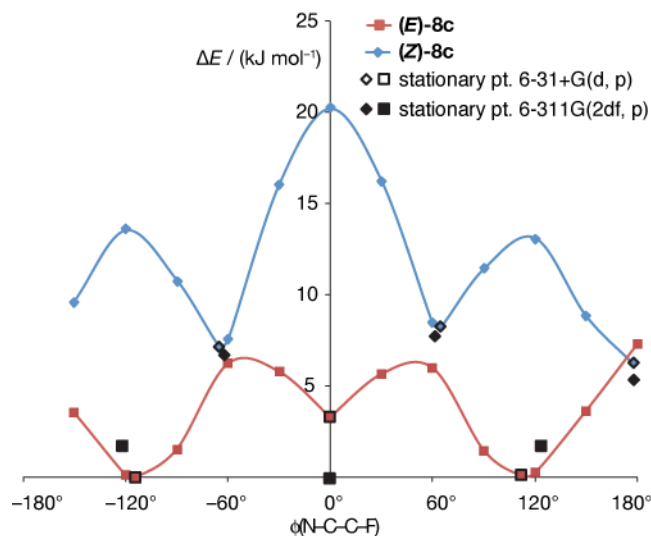


Figure 7-5. Bond rotation profiles of the quinoid β -fluoropyridoximine tautomers (*E*)- and (*Z*)-**8c**, calculated in water at the M06-2X/6-31+G(d, p) level. Data for both isomers are plotted relative to the lowest stationary point at this level of theory, (*-ac*)-(*E*)-**8c**. Full black symbols show conformer energies obtained with the larger 6-311G(2df, p) basis set, relative to lowest stationary point at that level, which is (*syn*)-(*E*)-**8c**.

The rotational profile for the less favoured (*Z*)-**8c** (Figure 7-5) is still similar to **8a, b** in that the minima are at $\phi_{\text{NCCF}} \approx \pm 60^\circ$ and 180° . However, the most stable conformer is now *anti*; and especially in water, the energy differences between the conformers are small ($< 3 \text{ kJ mol}^{-1}$). This is consistent with an electrostatic control (repulsion between N and F), which favours *anti*; in water, the electrostatic interaction is screened, so the differences between conformers are reduced. No stereoelectronic *gauche* preference is operational in this system. The barriers separating the minima are significantly lower in (*Z*)-**8c** than in **8a, b**, especially the central *-gauche* \rightarrow *+gauche* barrier, which is almost halved. The rotational barriers in (*Z*)-**8c** are not caused by *syn* vicinal interactions, but by the relatively weak 1,3-allylic (A1,3) strain between the fluoromethyl group and the vinylic CH (Figure 7-6).

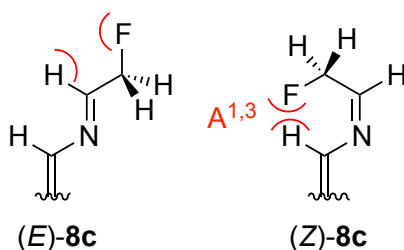


Figure 7-6. Unfavourable interactions responsible for the highest NCCF rotational barrier in (*E*)- and (*Z*)-**8c**, respectively.

In (*E*)-**8c**, finally, the NCCF rotation is almost unhindered, and the conformers are essentially energetically degenerate in both vacuum and water. Most notable is the interchange between minima and maxima in the torsional profile for (*E*)-**8c** as compared to all previous cases. The conformers of (*E*)-**8c** are at $\varphi_{\text{NCCF}} \approx \pm 120^\circ$ and 0° , while $\varphi_{\text{NCCF}} \approx \pm 60^\circ$ and 180° correspond to (low) barriers. The barriers are caused by unfavourable *syn* vicinal interactions (Figure 7-6). These are, however, much weaker than in, *e.g.*, **8a**, because of the wider sp^2 bond angle at C_α ($\text{N}-\text{C}_\alpha-\text{C}_\beta$ is 131° in **8c**, but 111° in **8a**), which positions the terminal atoms of the NCCF torsion further apart (2.93 \AA in **8c** vs. 2.57 \AA in **8a** at $\varphi_{\text{NCCF}} = 0^\circ$). The flatness of the potential energy surface for NCCF rotation in (*E*)-**8c** is also reflected in the relatively large influence of basis set and solvation on relative energies (several kJ mol^{-1}) and exact positions of the minima ($\sim 10^\circ$).

7.3.1.4 C–F activation in β -fluoropyridoximine

Relating back to Dunathan's hypothesis, it was finally analysed the effect of the NCCF torsion angle on the activation of the C–F bond in the quinoid (*E*)-**8c**. As a simple measure of C–F activation, the C–F bond length as a function of φ_{NCCF} is shown in Figure 7-7. The C–F bond is lengthened by more than 0.02 \AA at $\varphi_{\text{NCCF}} = \pm 90^\circ$, when it is exactly antiperiplanar to the conjugated imine system, compared to the *syn* conformer, when it is co-planar with the conjugated system. In the two *anticlinal* conformers ($\varphi_{\text{NCCF}} \approx \pm 120^\circ$), the C–F bond is almost as elongated as at $\varphi_{\text{NCCF}} = \pm 90^\circ$. The NBO analysis revealed that there is indeed substantial $\pi_{\text{N}=\text{C}} \rightarrow \sigma_{\text{C}-\text{F}}^*$ donation when the C–F bond is antiperiplanar to the π -system. The antibonding

$\sigma^*_{\text{C-F}}$ orbital has an occupancy of 0.05e, which weakens the C–F bond. In that conformation, the overlap between the donor and acceptor NBOs is optimal (Figure 7-8). With -38 kJ mol^{-1} , the energy gained from this donor–acceptor interaction is second only to the contributions from delocalization within the π -system itself. By contrast, in the *syn* conformer, the C–F bond is antiperiplanar to the $\text{C}\alpha\text{--H}$ bond, which results in a much weaker $\sigma_{\text{C-H}} \rightarrow \sigma^*_{\text{C-F}}$ interaction. The $\sigma^*_{\text{C-F}}$ orbital is occupied by only 0.01e, and the delocalization energy amounts to -15 kJ mol^{-1} . In the *E*-quinoid, therefore, the stereoelectronic situation is such that the C–F bond (or any other acceptor bond) is preferentially activated when it is antiperiplanar to the extended π -system.

Moreover, this conformation is accessible at very little energetic cost as the NCCF rotation is practically unhindered.

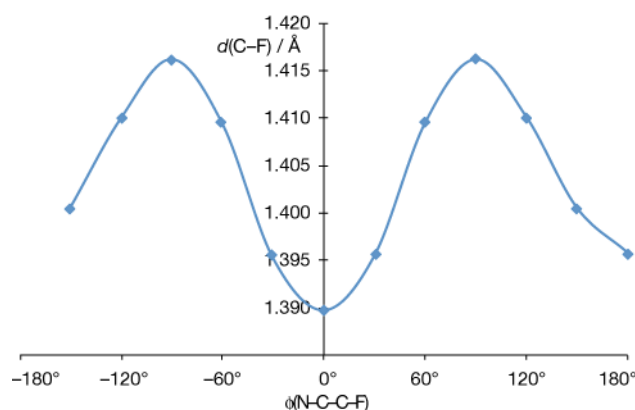


Figure 7-7. Variation of the C–F bond length with NCCF torsion in (*E*)-**8c**, calculated using M06-2X/6-31+G(d, p) in IEFPCM continuum water.

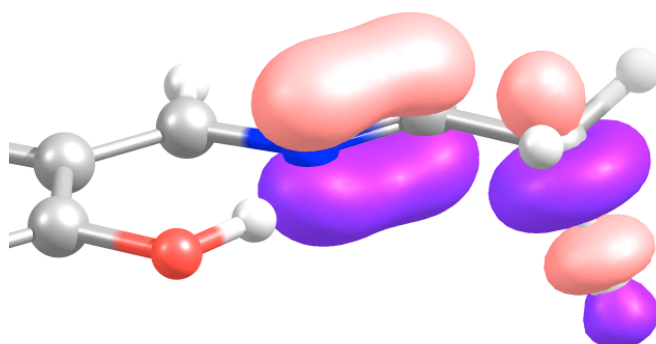


Figure 7-8. Illustration of the stereoelectronic C–F activation in (*E*)-**8c** at $\phi_{\text{NCCF}} = 90^\circ$. Shown are the occupied π_{NC} and the vacant σ^*_{CF} NBOs.

7.4 Conclusions

The main objective of this chapter was to clarify if the β - fluoroamines adopt *gauche* conformation as other fluorinated compounds containing an electron-withdrawing group in the β -position.

The calculations were performed at the M06-2X/6-31+G(d, p) level in IEFPCM continuum water or gas phase. NBO analysis was conducted to investigate the stereoelectronic effects in the β - fluorinated compounds.

In conclusion, we have presented theoretical and crystallographic evidence that β - fluoroamines preferentially adopt one of two possible *gauche* conformations, due to hyperconjugative donor-acceptor interaction, and that generally the overall order of stability is $-gauche < anti < +gauche$. The preference of the $-gauche$ conformer over the $+gauche$ conformer is likely due to electrostatic repulsion and constitutes a rare example of an *anti* conformer of a β -fluoroamine derivative being more energetically favourable to the *gauche*. Donor-acceptor interaction energies were calculated and the established the preference of the $-gauche$ over the *anti* conformer. The *gauche* effect is more pronounced in β - fluoroamines with electron-withdrawing substituents. The conformational analysis of β - fluoropyridoxamines (**8**) revealed that the $-gauche$ conformer is the most stable for tautomers **8a** and **8b**. The $+gauche$ is stabilized over the *anti* conformer and the order of stability has changed to $-gauche < +gauche < anti$. The (*E*)-**8c** has higher stability than the (*Z*)-**8c** with minima at $\varphi_{\text{NCCF}} \approx \pm 120^\circ$ and 0° .

Computational methods have enabled us to compare and contrast the conformational preferences of both the “internal” (F–C–C–N=C) aldimine and “external” (F–C–C=N–C) quinoid forms of the transient intermediates which are implicated in the mechanism of PLP-dependent enzyme inhibition by β - fluoroamines. A stereoelectronic rational for this conformational preference is a stabilising hyperconjugative interaction from the $\sigma_{\text{C–H}}$ bond into the low lying σ^* orbital of the C–F bond ($\sigma_{\text{C–H}} \rightarrow \sigma^*_{\text{C–F}}$).¹⁷⁰

References

- (1) Faraci, W. S.; Walsh, C. T. *Biochemistry* **1989**, 28, 431.
- (2) Koch, W.; Holthausen, M. C. *A chemist's guide to Density Functional Theory*; second ed.; Wiley & Sons, Ltd., 2001.
- (3) Born, M.; Oppenheimer, R. *Ann. Phys.-Berlin* **1927**, 84, 457.
- (4) Slater, J. C. *Phys. Rev.* **1930**, 35, 210.
- (5) Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, 136, B864.
- (6) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, 140, 1133.
- (7) Leach, A. R. *Molecular Modelling Principles and Applications*; second ed.; Pearson, 2001.
- (8) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, 58, 1200.
- (9) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785.
- (10) Filatov, M.; Thiel, W. *Mol. Phys.* **1997**, 91, 847.
- (11) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, 45, 13244.
- (12) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, 77, 3865.
- (13) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, 110, 13126.
- (14) Staroverov, V. N.; Scuseria, G. E.; Tao, J. M.; Perdew, J. P. *J. Chem. Phys.* **2003**, 119, 12129.
- (15) Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, 52, 997.
- (16) Marques, M. A. L.; Gross, E. K. U. *Annu. Rev. Phys. Chem.* **2004**, 55, 427.
- (17) Hückel, E. *Z. Phys.* **1931**, 70, 204.
- (18) Cramer, C. In *Essentials of Computational Chemistry*; Wiley & Sons, Ltd.: 2004.
- (19) Dewar, M. J. S.; Thiel, W. *Theor. Chim. Acta* **1977**, 46, 89.
- (20) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, 107, 3902.
- (21) Mordasini, T. Z.; Thiel, W. *Chimia* **1998**, 52, 288.
- (22) Ryde, U. *J. Comput-Aided. Mol. Des.* **1996**, 10, 153.
- (23) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, 100, 10580.
- (24) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, 11, 700.
- (25) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, 103, 227.
- (26) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, 102, 3586.
- (27) Thery, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, 15, 269.
- (28) Gao, J. L.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, 102, 4714.
- (29) Schlick, T. *Molecular Modeling and Simulations*; Springer, 2000.
- (30) Verlet, L. *Phys. Rev.* **1967**, 159, 98.

- (31) Hockney, R. W. *Methods Comput. Phys.* **1970**, 136.
- (32) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, 81, 3684.
- (33) Nosé, S. *Mol. Phys.* **1984**, 52, 255.
- (34) Hoover, W. G. *Phys. Rev. A* **1985**, 31, 1695.
- (35) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, 52, 7182.
- (36) Spoel, D. v. d.; Lindahl, E.; Hess, B.; Buuren, A. R. v.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; Drunen, R. v.; Berendsen, H. J. C. *Gromacs User Manual version 3.3*, 2005.
- (37) Senn, H. M. In *Lecture notes : Molecular Modelling*; University of Glasgow: Glasgow, 2008.
- (38) Anderson, J. L. R.; Chapman, S. K. *Mol. BioSyst.* **2006**, 2, 350.
- (39) Gribble, G. W. *Acc. Chem. Res.* **1998**, 31, 141.
- (40) Subbaraju, G. V.; Golakoti, T.; Patterson, G. M. L.; Moore, R. E. *J. Nat. Prod.* **1997**, 60, 302.
- (41) Gouda, H.; Matsuzaki, K.; Tanaka, H.; Hirono, S. *J. Am. Chem. Soc.* **1996**, 118, 13087.
- (42) Murphy, C. D. *J. Appl. Microbiol.* **2003**, 95, 203.
- (43) Faulkner, D. J.; Stallard, M. O.; Fayos, J.; Clardy, J. *J. Am. Chem. Soc.* **1973**, 95, 3413.
- (44) Deng, H.; O'Hagan, D.; Schaffrath, C. *Nat. Prod. Rep.* **2004**, 21, 773.
- (45) Neumann, C. S.; Fujimori, D. G.; Walsh, C. T. *Chem. Biol.* **2008**, 15, 99.
- (46) Blasiak, L. C.; Drennan, C. L. *Acc. Chem. Res.* **2009**, 42, 147.
- (47) Butler, A.; Sandy, M. *Nature* **2009**, 460, 848.
- (48) Chen, X.; Pée, K.-H. v. *Acta Biochim. Biophys Sin* **2008**, 40, 183.
- (49) Fujimori, D. G.; Walsh, C. T. *Curr. Opin. Chem. Biol.* **2007**, 11, 553.
- (50) van Pée, K. H.; Dong, C. J.; Flecks, S.; Naismith, J.; Patallo, E. P.; Wage, T. *Adv. Appl. Microbiol.* **2006**, 59, 127.
- (51) Wagner, C.; El Omari, M.; König, G. M. *J. Nat. Prod.* **2009**, 72, 540.
- (52) Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; Garneau-Tsodikova, S.; Walsh, C. T. *Chem. Rev.* **2006**, 106, 3364.
- (53) Winter, J. M.; Moore, B. S. *J. Biol. Chem.* **2009**, 284, 18577.
- (54) Vilter, H. *Phytochemistry* **1984**, 23, 1387.
- (55) Feiters, M. C.; Leblanc, C.; Kupper, F. C.; Meyer-Klaucke, W.; Michel, G.; Potin, P. *J. Am. Chem. Soc.* **2005**, 127, 15340.
- (56) Messerschmidt, A.; Wever, R. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, 93, 392.
- (57) Murphy, C. D. *Nat. Prod. Rep.* **2006**, 23, 147.
- (58) Dorrestein, P. C.; Yeh, E.; Garneau-Tsodikova, S.; Kelleher, N. L.; Walsh, C. T. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, 102, 13843.
- (59) Lin, S.; Van Lanen, S. G.; Shen, B. *J. Am. Chem. Soc.* **2007**, 129, 12432.
- (60) Nowak-Thompson, B.; Chaney, N.; Wing, J. S.; Gould, S. J.; Loper, J. E. *J. Bacteriol.* **1999**, 181, 2166.
- (61) Wynands, I.; van Pée, K. H. *FEMS Microbiol. Lett.* **2004**, 237, 363.
- (62) Bitto, E.; Huang, Y.; Bingman, C. A.; Singh, S.; Thorson, J. S.; Phillips, G. N. *Proteins Struct. Funct. Gen.* **2008**, 70, 289.
- (63) Zhu, X. F.; De Laurentis, W.; Leang, K.; Herrmann, J.; Lhiefeld, K.; van Pée, K. H.; Naismith, J. H. *J. Mol. Biol.* **2009**, 391, 74.

- (64) Dong, C. J.; Flecks, S.; Unversucht, S.; Haupt, C.; van Pée, K. H.; Naismith, J. H. *Science* **2005**, *309*, 2216.
- (65) Zehner, S.; Kotzsch, A.; Bister, B.; Sussmuth, R. D.; Mendez, C.; Salas, J. A.; van Pée, K. H. *Chem. Biol.* **2005**, *12*, 445.
- (66) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
- (67) Yeh, E.; Blasiak, L. C.; Koglin, A.; Drennan, C. L.; Walsh, C. T. *Biochemistry* **2007**, *46*, 1284.
- (68) Yeh, E.; Cole, L. J.; Barr, E. W.; Bollinger, J. M.; Ballou, D. P.; Walsh, C. T. *Biochemistry* **2006**, *45*, 7904.
- (69) Yeh, E.; Garneau, S.; Walsh, C. T. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 3960.
- (70) Chang, C. J.; Floss, H. G.; Hook, D. J.; Mabe, J. A.; Manni, P. E.; Martin, L. L.; Schroder, K.; Shieh, T. L. *J. Antibiot.* **1981**, *34*, 555.
- (71) Chen, X. P.; Huang, M. F.; Wang, B. *Chin. J. Chem.* **2008**, *26*, 1486.
- (72) Keller, S.; Wage, T.; Hohaus, K.; Holzer, M.; Eichhorn, E.; van Pee, K. H. *Angew. Chem. Int. Ed.* **2000**, *39*, 2300.
- (73) Ligon, J. M.; Hill, D. S.; Hammer, P. E.; Torkewitz, N. R.; Hofmann, D.; Kempf, H. J.; van Pee, K. H. *Pest Manage. Sci.* **2000**, *56*, 688.
- (74) van Pée, K. H.; Ligon, J. M. *Nat. Prod. Rep.* **2000**, *17*, 157.
- (75) Zhou, P.; Mocek, U.; Siesel, B.; Floss, H. G. *J. Basic Microbiol.* **1992**, *32*, 2009.
- (76) Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'Connor, S. E.; Walsh, C. T. *Nature* **2005**, *436*, 1191.
- (77) Wuosmaa, A. M.; Hager, L. P. *Science* **1990**, *249*, 160.
- (78) Carey, F.; Sundberg, F. *Advanced Organic Chemistry: Part A: Structure and Mechanisms: Structure and Mechanisms PartA: Structure and Mechanism*; fifth ed.; Springer, 2007.
- (79) *TURBOMOLE*; V. 5.9.1, 2008.
- (80) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
- (81) Häser, M.; Ahlrichs, R. *J. Comput. Chem.* **1989**, *10*, 104.
- (82) Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346.
- (83) von Arnim, M.; Ahlrichs, R. *J. Comput. Chem.* **1998**, *19*, 1746.
- (84) von Arnim, M.; Ahlrichs, R. *J. Chem. Phys.* **1999**, *111*, 9183.
- (85) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*.
- (86) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.
- (87) Klamt, A.; Schüürmann, G. *J. Chem. Soc. Perkin Trans. 2* **1993**, 799.
- (88) Li, J. B.; Zhu, T. H.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9.
- (89) *ChemShell*; 3. 4. dev., 2010.
- (90) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schafer, A.; Lennartz, C. *J. Mol. Struct. Theochem* **2003**, *632*, 1.
- (91) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177.
- (92) Liu, D. C.; Nocedal, J. *Math. Prog.* **1989**, *45*, 503.

- (93) Nocedal, J. *Math. Comput.* **1980**, 35, 773.
- (94) Baker, J. J. *J. Comput. Chem.* **1986**, 7, 385.
- (95) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. J. *J. Phys. Chem.* **1985**, 89, 52.
- (96) Deglmann, P.; Furche, F. *J. Chem. Phys.* **2002**, 117, 9535.
- (97) Deglmann, P.; Furche, F.; Ahlrichs, R. *Chem. Phys. Lett.* **2002**, 362, 511.
- (98) Francisco, J. S.; Sander, S. P. *Chem. Phys. Lett.* **1995**, 241, 33.
- (99) Cherney, D. P.; Duirk, S. E.; Tarr, J. C.; Collette, T. W. *Appl. Spectrosc.* **2006**, 60, 764.
- (100) Turner, A. J.; Moliner, V.; Williams, I. H. *Phys. Chem. Chem. Phys.* **1999**, 1, 1323.
- (101) Hammond, G. S. *J. Am. Chem. Soc.* **1955**, 77, 334.
- (102) Hamilton, C., University of Glasgow, 2008.
- (103) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96, 400.
- (104) Vriend, G. *J. Mol. Graphics* **1990**, 8, 52.
- (105) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins Struct. Funct. Bioinf.* **2005**, 61, 704.
- (106) Powers, N.; Jensen, J. H. *J. Biomol. NMR* **2006**, 35, 39.
- (107) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. *Biophys. J.* **2002**, 83, 1731.
- (108) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. *Nucleic Acids Res.* **2005**, 33, W368.
- (109) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. *Comput. Phys. Commun.* **1995**, 91, 43.
- (110) Lindahl, E.; Hess, B.; Van Der Spoel, D. *J. Mol. Model.* **2001**, 7, 306.
- (111) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, 26, 1701.
- (112) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. J. *J. Comput. Chem.* **2003**, 24, 1999.
- (113) <http://www.ks.uiuc.edu/>.
- (114) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, 25, 247.
- (115) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, 26, 1668.
- (116) *AMBER 10*, 2008.
- (117) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, 91, 1.
- (118) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2005**, 26, 114.
- (119) Pomelli, C. S.; Tomasi, J.; Barone, V. *Theor. Chem. Acc.* **2001**, 105, 446.
- (120) Tomasi, J.; Mennucci, B.; Cances, E. *J. Mol. Struct. Theochem* **1999**, 464, 211.
- (121) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, 97, 10269.

- (122) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (123) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
- (124) *Gaussian 03*; rev. E.01, 2004.
- (125) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.
- (126) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221.
- (127) Smith, W.; Forester, T. R. *J. Mol. Graphics* **1996**, *14*, 136.
- (128) Ristila, M.; Matxain, J. M.; Strid, A.; Eriksson, L. A. *J. Phys. Chem. B* **2006**, *110*, 16774.
- (129) Eliot, A. C.; Kirsch, J. F. *Annu. Rev. Biochem.* **2004**, *73*, 383.
- (130) Metzler, C. M.; Harris, A. G.; Metzler, D. E. *Biochemistry* **1988**, *27*, 4923.
- (131) *Gaussian 09*; rev. A.02, 2009.
- (132) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
- (133) Cortijo, M.; Llor, J.; Sanchezruiz, J. M. *J. Biol. Chem.* **1988**, *263*, 17960.
- (134) Diaz, I.; Montero, L. A.; Munoz, F.; Donoso, J. *Theochem-J. Mol. Struc.* **1998**, *433*, 279.
- (135) Echevarria, G. R.; Catalan, J.; Blanco, F. G. *Photochem. Photobiol.* **1997**, *66*, 810.
- (136) Harris, C. M.; Johnson, R. J.; Metzler, D. E. *Biochim. Biophys. Acta* **1976**, *421*, 181.
- (137) Matsushi, Y.; Martell, A. E. *J. Am. Chem. Soc.* **1967**, *89*, 1322.
- (138) Ledbetter, J. W.; Hnckel, J. M.; Cornish, T. J. *Photochem. Photobiol.* **1981**, *34*, 115.
- (139) Phillips, R. S.; Demidkina, T. V.; Zakomirdina, L. N.; Bruno, S.; Ronda, L.; Mozzarelli, A. *J. Biol. Chem.* **2002**, *277*, 21592.
- (140) Metzler, C. M.; Metzler, D. E. *Anal. Biochem.* **1987**, *166*, 313.
- (141) Dunathan, H. C. *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *55*, 712.
- (142) Bugg, T. D. H. *Introduction to Enzyme and Coenzyme Chemistry*; second ed.; Blackwell Publishing, 2004.
- (143) Hayashi, H. *J. Biochem.* **1995**, *118*, 463.
- (144) Hayashi, H.; Wada, H.; Yoshimura, T.; Esaki, N.; Soda, K. *Annu. Rev. Biochem.* **1990**, *59*, 87.
- (145) Vederas, J. C.; Floss, H. G. *Acc. Chem. Res.* **1980**, *13*, 455.
- (146) Berkowitz, D. B.; de la Salud-Bea, R.; Jahng, W. J. *Org. Lett.* **2004**, *6*, 1821.
- (147) Berkowitz, D. B.; Karukurichi, K. R.; de la Salud-Bea, R.; Nelson, D. L.; McCune, C. D. *J. Fluorine Chem.* **2008**, *129*, 731.
- (148) Lippert, B.; Metcalf, B. W.; Resvick, R. J. *Biochem. Biophys. Res. Commun.* **1982**, *108*, 146.
- (149) Walsh, C. T. *Annu. Rev. Biochem.* **1984**, *53*, 493.
- (150) Murali, D.; Flores, L. G.; Roberts, A. D.; Nickles, R. J.; DeJesus, O. T. *Appl. Radiat. Isot.* **2003**, *59*, 237.
- (151) Palfreyman, M. G.; Danzin, C.; Bey, P.; Jung, M. J.; Ribereaugayon, G.; Aubry, M.; Vevert, J. P.; Sjoerdsma, A. *J. Neurochem.* **1978**, *31*, 927.
- (152) Casella, L.; Gullotti, M. *J. Am. Chem. Soc.* **1983**, *105*, 803.
- (153) Nagata, H.; Doi, M.; Inoue, M.; Ishida, T.; Kamigauchi, M.; Sugiura, M.; Wakahara, A. *J. Chem. Soc. Perkin Trans. 2* **1994**, 983.

- (154) Tsai, M. D.; Weintraub, H. J. R.; Byrn, S. R.; Chang, C.; Floss, H. G. *Biochemistry* **1978**, *17*, 3183.
- (155) Béguin, C.; Hamman, S. *Org. Magn. Resonance* **1981**, *16*, 129.
- (156) Salon, M. C.; Hamman, S.; Béguin, C. G. *Org. Magn. Resonance* **1983**, *21*, 265.
- (157) Salon, M. C.; Hamman, S.; Béguin, C. G. *J. Fluorine Chem.* **1985**, *27*, 361.
- (158) Unver, H.; Kendi, E.; Guven, K.; Durlu, T. N. *Naturforsch. B* **2002**, *57*, 685.
- (159) Briggs, C. R. S.; Allen, M. J.; O'Hagan, D.; Tozer, D. J.; Slawin, A. M. Z.; Goeta, A. E.; Howard, J. A. K. *Org. Biomol. Chem.* **2004**, *2*, 732.
- (160) Briggs, C. R. S.; O'Hagan, D.; Howard, J. A. K.; Yufit, D. S. *J. Fluorine Chem.* **2003**, *119*, 9.
- (161) Bucher, C.; Sparr, C.; Schweizer, W. B.; Gilmour, R. *Chem.-Eur. J.* **2009**, *15*, 7637.
- (162) Buissonneaud, D. Y.; van Mourik, T.; O'Hagan, D. *Tetrahedron* **2010**, *66*, 2196.
- (163) Gooseman, N. E. J.; O'Hagan, D.; Peach, M. J. G.; Slawin, A. M. Z.; Tozer, D. J.; Young, R. J. *Angew. Chem. Int. Ed.* **2007**, *46*, 5904.
- (164) Gooseman, N. E. J.; O'Hagan, D.; Slawin, A. M. Z.; Teale, A. M.; Tozer, D. J.; Young, R. J. *Chem. Commun.* **2006**, 3190.
- (165) Mondelli, C.; Bucher, C.; Baiker, A.; Gilmour, R. *J. Mol. Catal. a-Chem.* **2010**, *327*, 87.
- (166) O'Hagan, D. *Chem. Soc. Rev.* **2008**, *37*, 308.
- (167) O'Hagan, D.; Bilton, C.; Howard, J. A. K.; Knight, L.; Tozer, D. J. *J. Chem. Soc. Perkin Trans. 2* **2000**, 605.
- (168) Sparr, C.; Schweizer, W. B.; Senn, H. M.; Gilmour, R. *Angew. Chem. Int. Ed.* **2009**, *48*, 3065.
- (169) Sparr, C.; Salamanova, E.; Schweizer, W. B.; Senn, H. M.; Gilmour, R. *Chem.-Eur. J.* **2011**, *17*, 8850.
- (170) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.
- (171) Goodman, L.; Gu, H. B.; Pophristic, V. *J. Phys. Chem. A* **2005**, *109*, 1223.
- (172) Weinhold, F.; Landis, C. R. *Valency and Bonding*; Cambridge university Press: Cambridge, 2005.

Appendix A

ZTRP ffAMBER atom types:

Num.at.	type	res.num	residue	atom	cgnr	charge	mass
1	amber99_45	1	ZTRP	OC1	1	-0.729189	16.0000
2	amber99_2	1	ZTRP	C	2	0.684283	12.0100
3	amber99_45	1	ZTRP	OC2	3	-0.729189	16.0000
4	amber99_11	1	ZTRP	CA	4	0.243709	12.0100
5	amber99_39	1	ZTRP	N	5	-0.546119	14.0100
6	amber99_17	1	ZTRP	H1	6	0.348930	1.0080
7	amber99_17	1	ZTRP	H2	7	0.348930	1.0080
8	amber99_17	1	ZTRP	H3	8	0.348930	1.0080
9	amber99_28	1	ZTRP	HA	9	0.027895	1.0080
10	amber99_11	1	ZTRP	CB	10	-0.036443	12.0100
11	amber99_18	1	ZTRP	HB1	11	0.038617	1.0080
12	amber99_18	1	ZTRP	HB2	12	0.038617	1.0080
13	amber99_14	1	ZTRP	CG	13	-0.227844	12.0100
14	amber99_13	1	ZTRP	CD1	14	-0.135691	12.0100
15	amber99_23	1	ZTRP	HD1	15	0.178769	1.0080
16	amber99_35	1	ZTRP	NE1	16	-0.381451	14.0100
17	amber99_17	1	ZTRP	HE1	17	0.376731	1.0080
18	amber99_8	1	ZTRP	CE2	18	0.223291	12.0100
19	amber99_3	1	ZTRP	CZ2	19	-0.326392	12.0100
20	amber99_22	1	ZTRP	HZ2	20	0.167964	1.0080
21	amber99_3	1	ZTRP	CH2	21	-0.102416	12.0100
22	amber99_22	1	ZTRP	HH2	22	0.127504	1.0080
23	amber99_3	1	ZTRP	CZ3	23	-0.228547	12.0100
24	amber99_22	1	ZTRP	HZ3	24	0.142845	1.0080
25	amber99_3	1	ZTRP	CE3	25	-0.170131	12.0100
26	amber99_22	1	ZTRP	HE3	26	0.143894	1.0080
27	amber99_4	1	ZTRP	CD2	27	0.172503	12.0100

FADHOOH using GAFF atom types:

Num.at.	type	res.num	residue	atom	cgnr	charge	mass
1	o	1	FADHOOH	O4	1	-0.58582	16.000000
2	c	1	FADHOOH	C4	2	0.75706	12.000000
3	n	1	FADHOOH	N3	3	-0.65460	14.000000
4	hn	1	FADHOOH	HN3	4	0.37347	1.000000
5	c	1	FADHOOH	C2	5	0.89381	12.000000
6	o	1	FADHOOH	O2	6	-0.61231	16.000000
7	ne	1	FADHOOH	N1	7	-0.69552	14.000000
8	c2	1	FADHOOH	C10	8	0.44607	12.000000

9	c3	1	FADHOOH	C4X	9	0.11535	12.000000
10	os	1	FADHOOH	O16	10	-0.23865	16.000000
11	oh	1	FADHOOH	O17	11	-0.36889	16.000000
12	ho	1	FADHOOH	H33	12	0.41009	1.000000
13	nh	1	FADHOOH	N5	13	-0.69142	14.000000
14	hn	1	FADHOOH	H32	14	0.38408	1.000000
15	ca	1	FADHOOH	C5X	15	0.40181	12.000000
16	ca	1	FADHOOH	C6	16	-0.42650	12.000000
17	ha	1	FADHOOH	H6	17	0.20361	1.000000
18	ca	1	FADHOOH	C7	18	0.15290	12.000000
19	c3	1	FADHOOH	C7M	19	-0.32778	12.000000
20	hc	1	FADHOOH	HM71	20	0.09974	1.000000
21	hc	1	FADHOOH	HM72	21	0.09974	1.000000
22	hc	1	FADHOOH	HM73	22	0.09974	1.000000
23	ca	1	FADHOOH	C8	23	0.15986	12.000000
24	c3	1	FADHOOH	C8M	24	-0.33026	12.000000
25	hc	1	FADHOOH	HM81	25	0.09868	1.000000
26	hc	1	FADHOOH	HM82	26	0.09868	1.000000
27	hc	1	FADHOOH	HM83	27	0.09868	1.000000
28	ca	1	FADHOOH	C9	28	-0.43240	12.000000
29	ha	1	FADHOOH	H9	29	0.27594	1.000000
30	ca	1	FADHOOH	C9A	30	0.01550	12.000000
31	nh	1	FADHOOH	N10	31	-0.02728	14.000000
32	c3	1	FADHOOH	C1'	32	0.01925	12.000000
33	h1	1	FADHOOH	H1'2	33	0.04684	1.000000
34	h1	1	FADHOOH	H1'1	34	0.04684	1.000000
35	c3	1	FADHOOH	C2'	35	0.29353	12.000000
36	oh	1	FADHOOH	O2'	36	-0.59109	16.000000
37	ho	1	FADHOOH	HO2'	37	0.39800	1.000000
38	h1	1	FADHOOH	H2'	38	0.02076	1.000000
39	c3	1	FADHOOH	C3'	39	0.18117	12.000000
40	oh	1	FADHOOH	O3'	40	-0.69746	16.000000
41	ho	1	FADHOOH	HO3'	41	0.46467	1.000000
42	h1	1	FADHOOH	H3'	42	0.01169	1.000000
43	c3	1	FADHOOH	C4'	43	0.15934	12.000000
44	oh	1	FADHOOH	O4'	44	-0.63517	16.000000
45	ho	1	FADHOOH	HO4'	45	0.38254	1.000000
46	h1	1	FADHOOH	H4'	46	0.04012	1.000000
47	c3	1	FADHOOH	C5'	47	0.11793	12.000000
48	h1	1	FADHOOH	H5'2	48	0.00735	1.000000
49	h1	1	FADHOOH	H5'1	49	0.00735	1.000000
50	os	1	FADHOOH	O5'	50	-0.32891	16.000000
51	p5	1	FADHOOH	P	51	1.19449	31.000000
52	o	1	FADHOOH	O1P	52	-0.82917	16.000000
53	o	1	FADHOOH	O2P	53	-0.82917	16.000000
54	os	1	FADHOOH	O3P	54	-0.51688	16.000000
55	p5	1	FADHOOH	PA	55	1.11022	31.000000
56	o	1	FADHOOH	O1A	56	-0.78639	16.000000
57	o	1	FADHOOH	O2A	57	-0.78639	16.000000
58	os	1	FADHOOH	O5B	58	-0.41386	16.000000

59	c3	1	FADHOOH	C5B	59	0.05748	12.000000
60	h1	1	FADHOOH	H52A	60	0.04425	1.000000
61	h1	1	FADHOOH	H51A	61	0.04425	1.000000
62	c3	1	FADHOOH	C4B	62	0.25931	12.000000
63	c3	1	FADHOOH	C3B	63	0.13971	12.000000
64	oh	1	FADHOOH	O3B	64	-0.67039	16.000000
65	ho	1	FADHOOH	HO3A	65	0.43534	1.000000
66	c3	1	FADHOOH	C2B	66	0.06537	12.000000
67	oh	1	FADHOOH	O2B	67	-0.67648	16.000000
68	ho	1	FADHOOH	HO2A	68	0.46703	1.000000
69	h1	1	FADHOOH	H2B	69	0.06306	1.000000
70	h1	1	FADHOOH	H3B	70	0.05920	1.000000
71	h1	1	FADHOOH	H4B	71	0.10380	1.000000
72	os	1	FADHOOH	O4B	72	-0.57471	16.000000
73	c3	1	FADHOOH	C1B	73	0.53616	12.000000
74	h2	1	FADHOOH	H1B	74	0.05847	1.000000
75	na	1	FADHOOH	N9A	75	-0.28686	14.000000
76	cc	1	FADHOOH	C8A	76	0.32311	12.000000
77	h5	1	FADHOOH	H8A	77	0.09125	1.000000
78	nd	1	FADHOOH	N7A	78	-0.62735	14.000000
79	ca	1	FADHOOH	C5A	79	-0.01912	12.000000
80	ca	1	FADHOOH	C6A	80	0.67209	12.000000
81	nh	1	FADHOOH	N6A	81	-0.79611	14.000000
82	hn	1	FADHOOH	H62A	82	0.37692	1.000000
83	hn	1	FADHOOH	H61A	83	0.37692	1.000000
84	nb	1	FADHOOH	N1A	84	-0.79703	14.000000
85	ca	1	FADHOOH	C2A	85	0.58578	12.000000
86	h5	1	FADHOOH	H2A	86	0.03325	1.000000
87	nb	1	FADHOOH	N3A	87	-0.78523	14.000000
88	ca	1	FADHOOH	C4A	88	0.55955	12.000000

Appendix B

pKa values of PrnA's residues

RESIDUE		PropKa pKa	MCCE pKa	H++ pKa
ARG	26A	11.95	>15	13.01
ARG	44A	13.32	>15	16.73
ARG	68A	12.36	13.32	12.41
ARG	88A	11.25	>15	14.67
ARG	95A	12.22	13.83	12.76
ARG	120A	11.54	>15	10.58
ARG	122A	12.36	13.15	12.58
ARG	154A	12.46	>15	15.04
ARG	173A	11.8	13.32	12.52
ARG	178A	11.59	13.87	14.48
ARG	182A	11.87	>15	13.76
ARG	191A	12.08	13.43	13.06
ARG	195A	12.29	13.40	12.97
ARG	207A	12.01	13.84	15.52
ARG	221A	11.93	12.80	12.11
ARG	256A	12.08	>15	14.74
ARG	281A	12.15	13.24	14.25
ARG	295A	11.45	13.74	14.12
ARG	320A	12.22	>15	13.07
ARG	323A	13.18	>15	18.11
ARG	326A	11.94	>15	16.24
ARG	372A	11.52	>15	17.92
ARG	389A	14.5	>15	24.34
ARG	401A	12.36	>15	19.75
ARG	411A	12.08	13.32	13.31
ARG	415A	12.08	>15	15.90
ARG	426A	12.42	13.56	17.61
ARG	473A	12.01	13.39	12.98
ARG	481A	12.73	12.94	15.46
ARG	496A	12.36	13.07	12.29
ARG	497A	12.01	12.52	12.17
ARG	501A	12.01	13.85	12.56
ARG	503A	12.15	13.86	13.86
ARG	514A	12.01	>15	15.32
ARG	517A	12.22	13.08	13.10
ARG	26B	11.95	>15	12.97
ARG	44B	13.32	>15	16.69
ARG	68B	12.36	13.53	12.45
ARG	88B	11.25	>15	14.67
ARG	95B	12.22	13.77	12.75
ARG	120B	11.54	>15	10.38
ARG	122B	12.36	13.28	12.61
ARG	154B	12.46	>15	15.03
ARG	173B	11.8	13.36	12.47

ARG	178B	11.59	13.70	14.45
ARG	182B	11.87	>15	13.79
ARG	191B	12.08	13.44	13.33
ARG	195B	12.29	13.47	12.40
ARG	207B	12.01	13.86	15.55
ARG	221B	11.93	12.94	12.10
ARG	256B	12.08	>15	14.79
ARG	281B	12.15	13.36	14.19
ARG	295B	11.45	13.69	14.15
ARG	320B	12.22	>15	13.03
ARG	323B	13.18	>15	17.90
ARG	326B	11.94	>15	16.27
ARG	372B	11.52	>15	17.96
ARG	389B	14.5	>15	24.35
ARG	401B	12.36	>15	19.93
ARG	411B	12.08	13.32	13.32
ARG	415B	12.08	>15	15.89
ARG	426B	12.42	>15	17.68
ARG	473B	12.01	>15	13.09
ARG	481B	12.73	13.54	15.24
ARG	496B	12.36	12.28	12.42
ARG	497B	12.01	12.62	12.18
ARG	501B	12.01	13.94	12.73
ARG	503B	12.15	13.96	13.83
ARG	514B	12.01	>15	14.61
ARG	517B	12.22	13.17	13.09
ASP	61A	1.96	1.29	0.99
ASP	92A	2.17	1.69	1.88
ASP	96A	3.35	1.98	1.35
ASP	97A	4.2	0.72	2.77
ASP	110A	0.81	1.02	-2.26
ASP	142A	1.56	2.98	-1.36
ASP	151A	2.22	3.70	1.79
ASP	163A	3.92	0.86	2.26
ASP	169A	1.15	2.60	0.40
ASP	185A	1.45	<0	-2.32
ASP	189A	3.94	3.38	3.77
ASP	212A	1.56	1.89	-0.35
ASP	216A	3.76	<0	-7.59
ASP	235A	2.87	2.49	2.82
ASP	238A	2.64	1.99	2.78
ASP	243A	2.2	2.18	-2.09
ASP	253A	2.98	4.09	3.50
ASP	254A	2.91	0.65	0.20
ASP	257A	3.77	3.68	3.56
ASP	296A	3.87	2.75	3.88
ASP	301A	4.15	3.36	3.68
ASP	310A	3.8	3.49	4.18
ASP	366A	1.98	2.06	-4.66
ASP	370A	1.25	1.00	-2.14

ASP	375A	4.05	3.91	3.87
ASP	386A	-1.21	1.46	-2.95
ASP	387A	5.46	0.85	-7.50
ASP	390A	1.67	<0	-6.04
ASP	402A	3.53	3.03	3.75
ASP	403A	-2	<0	-0.63
ASP	413A	4.15	3.21	3.73
ASP	418A	3.8	2.34	3.10
ASP	439A	1.92	1.77	1.28
ASP	440A	4.55	2.39	-1.03
ASP	448A	3.77	3.01	0.61
ASP	472A	3.42	3.11	1.06
ASP	511A	2.71	1.56	2.63
ASP	518A	3.48	4.34	4.24
ASP	61B	1.96	4.81	0.94
ASP	92B	2.17	0.96	1.89
ASP	96B	3.35	1.58	0.91
ASP	97B	4.2	1.80	2.53
ASP	110B	0.81	1.70	-2.39
ASP	142B	1.56	<0	3.20
ASP	151B	2.22	2.94	1.85
ASP	163B	3.92	4.03	2.31
ASP	169B	1.15	0.96	0.46
ASP	185B	1.45	2.49	-2.27
ASP	189B	3.94	<0	3.79
ASP	212B	1.56	3.41	-0.50
ASP	216B	3.76	2.07	-7.55
ASP	235B	2.87	<0	2.71
ASP	238B	2.64	2.40	2.71
ASP	243B	2.2	2.01	-2.14
ASP	253B	2.98	2.43	3.39
ASP	254B	2.91	4.04	0.18
ASP	257B	3.77	2.42	3.52
ASP	296B	3.87	3.65	3.83
ASP	301B	4.15	2.77	4.09
ASP	310B	3.8	3.03	4.26
ASP	366B	1.98	3.53	-0.02
ASP	370B	1.25	1.31	-3.45
ASP	375B	4.05	<0	4.17
ASP	386B	-1.21	3.72	-2.87
ASP	387B	5.46	1.68	-4.32
ASP	390B	1.67	<0	-6.78
ASP	402B	3.53	<0	3.73
ASP	403B	-2	3.08	-0.50
ASP	413B	4.15	<0	3.71
ASP	418B	3.8	3.27	3.10
ASP	439B	1.92	2.01	3.48
ASP	440B	4.55	1.19	-0.91
ASP	448B	3.77	1.93	0.16
ASP	472B	3.42	<0	1.15

ASP	511B	2.71	1.10	3.66
ASP	518B	3.48	1.54	4.03
CYS	109A	10.91	4.82	18.52
CYS	134A	11.56	>15	20.96
CYS	148A	6.29	>15	18.53
CYS	217A	12.59	>15	22.52
CYS	242A	13.67	>15	25.37
CYS	332A	12.67	>15	21.65
CYS	340A	11.77	>15	21.19
CYS	388A	12.86	>15	37.03
CYS	462A	8.62	>15	32.56
CYS	109B	10.91	>15	18.82
CYS	134B	11.56	>15	21.21
CYS	148B	6.29	>15	18.85
CYS	217B	12.59	>15	24.25
CYS	242B	13.67	>15	24.67
CYS	332B	12.67	>15	21.73
CYS	340B	11.77	>15	19.39
CYS	388B	12.86	>15	37.19
CYS	462B	8.62	>15	29.27
GLU	38A	4.44	1.09	-0.95
GLU	49A	6.07	1.91	-6.06
GLU	67A	5.56	5.22	-1.18
GLU	69A	4.78	4.02	4.42
GLU	123A	3.93	4.29	1.53
GLU	131A	4.88	1.12	-0.84
GLU	177A	4.5	4.05	3.89
GLU	186A	3.91	2.93	2.68
GLU	205A	4.1	4.54	4.14
GLU	210A	4.5	3.79	4.17
GLU	231A	2.25	<0	-1.61
GLU	260A	0.31	<0	-0.59
GLU	343A	7.33	6.48	4.05
GLU	346A	0.53	<0	-19.39
GLU	380A	3.63	3.22	-5.62
GLU	422A	1.65	2.02	0.22
GLU	445A	4.67	4.14	4.03
GLU	450A	4.84	7.28	4.34
GLU	483A	4.5	3.45	3.70
GLU	486A	4.64	4.15	4.48
GLU	489A	3.98	2.38	0.55
GLU	498A	4.86	4.68	4.43
GLU	500A	4.16	4.42	3.51
GLU	38B	4.44	1.16	-0.79
GLU	49B	5.4	<0	-6.41
GLU	67B	5.56	5.10	-1.10
GLU	69B	4.78	4.03	4.39
GLU	123B	3.93	4.28	1.60
GLU	131B	4.88	0.85	-1.51
GLU	177B	4.5	3.90	3.88

GLU	186B	3.91	3.15	2.67
GLU	205B	4.1	4.60	4.13
GLU	210B	4.5	3.76	4.18
GLU	231B	2.25	<0	-1.61
GLU	260B	0.31	<0	-1.38
GLU	343B	7.33	<0	5.03
GLU	346B	0.89	2.01	-19.38
GLU	380B	3.63	<0	-5.59
GLU	422B	1.65	<0	0.17
GLU	445B	4.67	3.96	3.99
GLU	450B	7.92	5.40	3.68
GLU	483B	4.5	3.08	3.68
GLU	486B	4.64	4.05	4.40
GLU	489B	3.98	2.24	0.48
GLU	498B	4.86	4.48	4.59
GLU	500B	4.16	4.43	3.78
HIS	98A	5.5	8.64	9.81
HIS	101A	6.6	<0	6.36
HIS	116A	1.21	<0	-6.30
HIS	158A	4.85	<0	5.84
HIS	161A	-0.19	<0	5.75
HIS	165A	2.61	<0	0.65
HIS	291A	6.5	7.50	7.17
HIS	363A	2.25	5.99	8.42
HIS	383A	3.06	5.95	6.05
HIS	395A	3.89	>15	16.85
HIS	412A	6.74	6.25	7.01
HIS	480A	3.1	6.33	4.65
HIS	98B	5.5	8.48	9.99
HIS	101B	6.38	<0	6.67
HIS	116B	1.21	<0	-6.71
HIS	158B	4.85	<0	6.15
HIS	161B	-0.19	<0	2.59
HIS	165B	2.61	<0	0.43
HIS	291B	6.5	7.51	7.19
HIS	363B	2.25	4.73	8.65
HIS	383B	3.06	5.47	6.12
HIS	395B	3.89	2.01	16.06
HIS	412B	6.74	6.23	7.02
HIS	480B	3.1	<0	4.52
LYS	3A	10.15	12.37	11.67
LYS	6A	10.5	10.57	10.20
LYS	57A	11.52	12.91	12.29
LYS	79A	5.48	8.30	16.94
LYS	83A	11.02	13.96	13.35
LYS	89A	10.5	10.75	10.83
LYS	121A	9.8	10.47	13.87
LYS	144A	11.39	12.43	12.77
LYS	172A	10.29	10.06	9.68
LYS	204A	10.5	11.35	10.90

LYS	230A	10.36	11.33	11.63
LYS	275A	11.15	13.38	16.22
LYS	304A	10.5	11.23	10.70
LYS	318A	10.5	10.75	10.30
LYS	325A	9.94	10.17	11.16
LYS	362A	9.29	11.81	14.81
LYS	421A	10.43	11.48	11.11
LYS	423A	9.94	11.13	13.51
LYS	428A	10.29	10.42	10.27
LYS	452A	10.83	>15	21.18
LYS	487A	10.29	11.50	12.78
LYS	3B	10.15	12.78	11.70
LYS	6B	10.5	10.55	10.21
LYS	57B	11.52	13.43	12.36
LYS	79B	5.48	8.27	15.62
LYS	83B	11.02	>15	13.55
LYS	89B	10.5	10.53	10.81
LYS	121B	9.8	10.33	13.65
LYS	144B	11.39	12.33	12.94
LYS	172B	10.29	10.26	9.69
LYS	204B	10.5	11.41	10.92
LYS	230B	10.36	11.19	11.63
LYS	275B	11.15	13.20	15.80
LYS	304B	10.5	11.28	10.75
LYS	318B	10.5	10.79	10.28
LYS	325B	9.94	10.28	11.07
LYS	362B	9.29	>15	14.36
LYS	421B	10.43	11.62	11.09
LYS	423B	9.94	13.01	12.62
LYS	428B	10.29	10.31	10.31
LYS	452B	10.83	>15	20.80
LYS	487B	10.29	11.71	12.85
N+	2A	7.93	8.62	
TYR	23A	12.53	>15	22.43
TYR	100A	19.7	>15	37.69
TYR	117A	14.42	>15	33.19
TYR	132A	14.82	>15	21.84
TYR	135A	16.61	>15	44.14
TYR	197A	9.88	10.59	11.80
TYR	239A	10.79	13.16	12.85
TYR	262A	14.58	>15	23.55
TYR	286A	11.6	>15	16.60
TYR	351A	14.52	>15	48.32
TYR	354A	12.81	>15	34.12
TYR	358A	15.94	>15	27.42
TYR	396A	20	>15	42.85
TYR	427A	12.47	>15	28.38
TYR	443A	15.32	>15	50.35
TYR	444A	16.81	>15	30.10
TYR	449A	11.54	13.90	28.05

TYR	460A	14.49	>15	32.92
TYR	461A	14.25	>15	37.32
TYR	510A	10.99	11.07	12.69
TYR	512A	12.38	>15	15.40
TYR	23B	12.53	>15	36.08
TYR	100B	19.7	>15	22.37
TYR	117B	14.42	>15	30.89
TYR	132B	14.82	>15	23.21
TYR	135B	16.61	>15	44.15
TYR	197B	9.88	10.48	12.94
TYR	239B	10.79	13.61	12.91
TYR	262B	14.58	>15	20.73
TYR	286B	11.6	>15	17.36
TYR	351B	14.52	>15	48.68
TYR	354B	12.81	>15	34.06
TYR	358B	15.94	>15	27.10
TYR	396B	20	>15	42.70
TYR	427B	12.47	>15	27.78
TYR	443B	16	>15	50.40
TYR	444B	15.71	>15	32.32
TYR	449B	11.54	>15	23.73
TYR	460B	14.49	>15	32.97
TYR	461B	14.25	>15	38.70
TYR	510B	10.99	11.14	12.94
TYR	512B	12.38	>15	15.69